

Dutch researchers' freedoms and responsibilities

Following damaging cuts in public spending last year, a new research policy in the Netherlands promises more freedom to researchers to set priorities. But national interests should not be forgotten.

By comparison with his predecessor, the Dutch science minister Loek Hermans is an unassuming character. He doesn't want his ministry to tell universities what research to pursue. In marked contrast, the previous science minister, Jo Ritzen, wanted to prescribe a university's number of PhD students. Although Hermans' job is to take responsibility for the state of research in the Netherlands, he is happy to let researchers have much more say in setting priorities. Ritzen, on the other hand, wanted to do the prioritizing.

For all Ritzen's determination to link research to socio-economic goals, in the end he seems to have stumbled on several counts. He failed to carry through a transfer of funds from the universities to the research funding agency, the Netherlands Organization for Scientific Research (NWO), in a bid to strengthen central prioritizing. And he failed to prevent significant cuts in Dutch research last year. All the signs are, now, that Dutch research is set for a more positive future thanks to Hermans' greater degree of political clout within the Dutch cabinet.

Last month, Hermans unveiled a plan that embodies his philosophy but which would also, if implemented, take universities a welcome step back from overly burdensome accountability. Under the terms spelt out in *Science Budget 2000*, ministerial micromanagement will be a thing of the past. Universities will submit strategic plans to the NWO every four years, rather than two-yearly reports of progress to the ministry. The NWO, in turn, will craft a national research plan based on universities' submissions. In order to foster awareness of potential economic and social relevance, foresight studies developed by the independent Advisory Council for Science and Technology Policy will be available for universities to take into account as they see fit.

These changes signal increased freedom for the Netherlands' 13 research universities to develop their own strategies, and increased clout to the NWO in developing national research and innovation. It also signals more flexibility in the allocation of funds to universities, hitherto — for the past 20 years — based rigidly on outdated statistics of student numbers.

The plan deserves two and a half cheers: one cheer for enhancing the role of universities in setting the priorities of their 2.4 billion guilders (US\$1.2 billion) research budget; another for decreasing the bureaucracy they have to submit to; and half a cheer for the success of Hermans in halting the funding decline — the document announces that funding cuts previously planned for the next few years will not occur. Research agencies and universities and, eventually (too eventually, perhaps), the ministry will shift some money into an innovation fund, set to grow to at least 75 million guilders, to develop fundamental research in priority areas such as bioinformatics and cognitive sciences.

It would be wrong to conclude that the Netherlands has drawn back from the idea that science should help develop the country economically and socially. Rather, the policy shift is an attempt to set a new balance of ministerial responsibility on the one hand and, on the other, to delegate authority to universities and the NWO, trusting them not only to pursue interesting questions but also to keep the national interest in mind. The Dutch research community packs a strong punch for one so small. The new balance of forces is a liberation that researchers and their institutions deserve, but their freedom will have to be tempered with the assumption of enhanced responsibility for considering long-term social and economic goals if they want to avoid a return to micromanagement. □

Big Science comes of age

Talk of the demise of Big Science is premature. But its characteristics have changed significantly.

When Alvin Weinberg coined the phrase 'Big Science' almost 40 years ago, the "monuments" to it that he listed — "huge rockets, high-energy accelerators, high-flux reactors" — were all identified with the physical sciences. To many, high-energy physics was the prototypical Big Science, and Weinberg himself — to the chagrin of many biologists — described such endeavours as "symbols of our time".

There have been times recently when the biologists have seemed keen to take over this mantle. In the early days of the Human Genome Project, for example, there was much talk of how this would put genetics into the Big Science league. Furthermore, US politicians are sending signals that their support for particle accelerators may be ending (see page 390) just at a time when the National Institutes of Health is making its first major contributions to the construction of synchrotron radiation facilities (see page 395). It is tempting to talk in terms of the swing of a pendulum, and to suggest that biology is now

beginning to enjoy the role that physics and space-based research have occupied for the past 50 years.

Such a characterization is misleading. Weinberg and those who picked up his Big Science idea were thinking of large experimental facilities that became the centrepiece of one or a few major collaborative research programmes. In contrast, the new synchrotrons and other devices are multi-user, multi-experimental facilities.

Furthermore, a key characteristic of the new facilities is that their use is not dominated by any one scientific discipline. The rapidly growing use of synchrotrons by structural biologists, or even of space missions by 'exobiologists', does not represent a take-over in any sense. Rather, both moves illustrate a more significant and long-lasting trend: the growing collaboration between separate scientific disciplines, and thus the emergence at new scales of a truly interdisciplinary approach to scientific problems. Big Science is not dead — it is merely coming of age. □



Japan declares five-year plan to double genome research funds

[TOKYO] The Japanese government is to step up support for its genomics industry, including a five-year plan to double spending on genome-related research and a plan to decode 30 per cent of human gene sequences by 2001.

The strategy was announced on 13 July by five science-related ministries (see *Nature* 400, 305; 1999). It will focus on analysing the human genome and creating a database of single-nucleotide polymorphisms (SNPs) in the Japanese population, which may lead to new drugs and diagnostic techniques.

The government also plans to create national centres for genomics research by 2001, and has pledged to expand its support for venture businesses. The latter will come through a scheme similar to the US programme Small Business Innovation Research, which funds small businesses carrying out innovative research with commercial potential.

The plan is in line with the government's 10-year programme for producing a 25-fold expansion in Japan's biotechnology market (see *Nature* 397, 554; 1999). That programme is aimed to create a market worth ¥25 trillion (US\$213 billion) and to help create 1,000 start-up companies by 2010.

The move also reflects a feeling among Japanese researchers and industry that Japan's biotechnology research, particularly its genome projects, is lagging severely behind those in the West (see *Nature* 399, 96; 1999).

For example, Japan's contribution to the Human Genome Project accounts for just 8 per cent of the total effort, and, according to the Ministry of International Trade and Industry (MITI), Japan's annual genomics spending is ¥560 billion, less than a quarter of that of the United States.

The government plans to inject ¥2 trillion over the next five years, bringing spending on genomics to at least 0.2 per cent of the country's gross domestic product.

Ministries involved include the Science and Technology Agency, MITI, the Ministry of Education, Science, Sports and Culture, the Ministry of Health and Welfare, and the Ministry of Agriculture, Forestry and Fisheries. The ministries will seek appropriations in this year's second supplementary budget, to be compiled in the autumn.

Japan's partners in the Human Genome Project are accelerating their sequencing effort in an attempt to sequence 10 per cent of the human genome in a 'working draft' by 2001. Japan is increasing its effort to develop technologies in functional genomics, particularly through the use of full-



Population genetics? Japan plans a database of the nation's single-nucleotide polymorphisms.

length human complementary DNA.

According to MITI, which oversees Japan's cDNA project (see *Nature* 398, 644; 1999), the aim is to have sequenced 30,000 cDNA clones by 2001. This would account for a third of expressed human genes. The project is led by Tokyo University's Institute of Medical Sciences and Japanese companies, with the goal of creating a central

repository of cDNA clones for medical and research applications.

These data will be used for the SNP programme, which aims to map 100,000–150,000 SNPs within two years. SNPs in coding sequences will be studied, as these are thought to be more likely to have functional significance than random SNPs.

The SNP programme would become the first genome project to be carried out on an interministerial level. In the past, such efforts have been criticized as being badly organized, with ministries carrying out individual genome projects.

The Science and Technology Council, Japan's principal science policy-making body, is calling for the creation of three national research centres focusing on applied genomics research, stem-cell research and plant genome research.

Although the government's biotechnology strategy promises some support for agrobiotechnology, the goal of sequencing the rice genome by 2008 remains unchanged.

"The Rice Genome Sequencing Project is pretty low on the priority list under the government's new biotechnology strategy," says an official from the agriculture ministry, which leads the project.

Asako Saegusa

UK ignored BSE vaccine advice, inquiry told

[LONDON] Warnings from scientists in the late 1980s that vaccines might transmit bovine spongiform encephalopathy (BSE) to humans were not passed on by the UK government or its safety committees, despite assurances to the contrary.

The news was revealed in evidence to the inquiry into Britain's outbreak of BSE, or 'mad cow disease'. The inquiry has been hearing from the scientists who first advised the government on the implications of the outbreak.

A working party headed by Sir Richard Southwood, former professor of zoology at the University of Oxford, faced questions about its approach to risk evaluation, management and communication.

Members of the working party said they did not act under government instructions and that their report had not been delayed to avoid embarrassment. The group sought to be "fiercely independent", said Southwood.

But he admitted that it had been cautious over the wording of its advice on the safety of vaccines, concerned that the public might boycott vaccination programmes.

The working party communicated its

concerns to the Committee for the Safety of Medicines and the Department of Health. But the committee seems to have disregarded the advice, and the department failed to pass warnings to pharmaceutical companies.

In contrast, the committee and government officials seem to have been reassured by the working party's use of the word 'remote' to describe the likely transmission of BSE to humans. But Southwood said this had not been intended by the authors.

The working party also pointed out that it considered the potential for transmission of BSE sufficiently serious to write to the Ministry of Agriculture, Fisheries and Food with interim advice after its initial meeting.

In a written statement to the inquiry, working party member Sir Anthony Epstein emphasized that "the existing basic scientific information was patchy", adding that "the sudden unexpected emergence of BSE, at a time when work in the field was minimal and experts few, provides a grim warning of the dangers of letting many aspects of basic science wither for lack of funds".

Natasha Loder

Senate baulks at cost of collider project

[WASHINGTON] The US Senate is moving to abandon the Next Linear Collider (NLC), a 30-kilometre long accelerator that high-energy physicists would like to build in the United States as their next major experiment after Europe's Large Hadron Collider.

The move follows a finding by the Department of Energy that the project could cost as much as \$7.9 billion to complete.

The Senate energy and water appropriations subcommittee, which monitors the department's budget, has called for NLC research costs to be cut from \$14 million to \$6 million next year. Senate staff say that the intention is to end funding altogether, but over two years, to minimize disruption at the Stanford Linear Accelerator Center (SLAC) in California, where most of the research is based.

The Senate move is motivated by a belief that the machine cannot be built in the foreseeable future, and that further research into its design is therefore pointless. One Senate staffer says that, even with large contributions from abroad, the NLC would cost the United States \$5 billion. "I cannot envisage any scenario in which the US will make that contribution," says the staffer.

Last month, after receiving a favourable review of the project together with the cost estimate, the energy department decided it was not ready to proceed with the conceptual



Cushioning the blow: Senators want to phase out linear collider research over two years to avoid too much disruption at the Stanford Linear Accelerator Center (left).

design phase that is required before any decision is taken to build the collider.

According to Martha Krebs, assistant secretary of energy, it would "look premature" to start the conceptual design, which would cost about \$70 million over two years. "If you wait for a year or two you might actually have a chance to convince people that this is a good project," she says.

The energy and water appropriations subcommittee in the House of Representatives supports Krebs' plan to continue NLC research. A conference between representatives of the House and Senate subcommittees in September will determine the immediate fate of the project.

But, even if collider research survives this year, the hesitancy of both the Congress and the administration to take the project further shows how difficult it will be for US physicists to build any machine with a price tag of much more than \$1 billion. The Senate move came despite the fact that Michael Witherell, the new director of Fermilab, outside Chicago, has buried the laboratory's rivalry with SLAC and sought to get involved in the NLC project.

The failure of even a unified high-energy physics community to generate momentum for the project confirms its waning influence, according to some observers in Washington. "There are a lot of other great ideas in science that don't cost billions of dollars," says one administration official.

"The American public doesn't care about staying at the forefront of high-energy physics." The official suggests that the end of the Cold War, together with declining public faith in the ability of physicists to come up with a clear explanation of the fundamental nature of matter, may help to explain reluctance to back the project.

Burt Richter, the director of SLAC, dismisses such arguments. He says that Congress initially approved the Superconducting Super Collider (SSC), at an estimated cost of \$6 billion, after the Cold War had ended.

Richter says the cost estimate for the NLC is highly conservative, to ensure that it will not drift up later. "The arguments now are not very different from the ones over the SSC," he says. "Every place in the world agrees that the next machine needs to be a big linear collider."

Witherell says: "What's needed is for the research to continue, so that we can work on things that will bring the cost down."

The NLC would accelerate electrons to 500 GeV, with the possibility of an upgrade that would take them to 1 TeV. The Stanford Linear Collider, the most powerful built so far, took electrons to 100 GeV. Scientists at KEK in Japan and at DESY in Germany are also researching colliders comparable to the NLC. Physicists hope for a global collaboration to build one machine, with the host bearing two-thirds of the cost.

Colin Macilwain

Neurogenetic institute comes to California

[SAN DIEGO] A Neurogenetic Institute with at least 30 faculty posts is to be set up at the University of Southern California (USC) with part of a \$110 million grant to the School of Medicine.

The grant, from the W. M. Keck Foundation, is one of the largest ever given to a US medical college. It is due to be announced today (29 July) along with plans to name the university's medical school after the foundation.

The Neurogenetic Institute will combine the efforts of 50 faculty members from within USC with the 30 to 35 new appointments. It will be housed in a research centre to be completed in two years on USC's Health Science Campus in East Los Angeles. USC officials estimate that one-third of the Keck grant will be used for the neurogenetic initiative.

Brian Henderson, a medical epidemiologist, will direct the institute, which will use an interdisciplinary approach to examine diseases such as Alzheimer's, Parkinson's, glaucoma, retinitis pigmentosa, macular degeneration, schizophrenia and manic depression.

"The challenge is to unravel the variables leading to neurological disorders and to

obtain data on the malfunction of the expressed genes that cause disease," says Henderson, the former director of the USC/Kenneth Norris Jr Comprehensive Cancer Center.

Principal research goals for the institute will include sequencing and characterizing critical nervous system genes, characterizing genotypic variations, and determining their role in disorders of sensory, motor and cognitive functions. Other important goals include identifying the genetic mechanisms in the molecular pathogenesis of neuropsychiatric and neurological disorders to devise diagnostics and therapeutics, as well as linking genetic susceptibilities with environmental and psychological risk factors.

Henderson says the USC medical school has historically been oriented primarily towards clinical care, but is now seeking to enhance its research efforts. "What USC badly needs is research faculty," he says.

As a condition of the Keck grant, USC has agreed to raise an additional \$330 million for its medical school. Established in 1954 by the founder of Superior Oil, the Keck foundation of Los Angeles has previously given more than \$140 million to USC and its faculty members.

Rex Dalton

NIH budget prospect bright despite political stalemate

[WASHINGTON] Divided Republicans in the US House of Representatives last week put off until September further efforts to craft a bill to fund the National Institutes of Health (NIH) in fiscal year 2000, which starts on 1 October. They had been forced to concede that broader budget politics have thrown up obstacles that are currently insurmountable.

But the delay does not necessarily portend ill for the biomedical agency. As efforts to agree on a broad funding bill to include the NIH were failing last week, Congressman John Porter (Republican, Illinois), chairman of the labour, health and human services and education subcommittee of the House Appropriations Committee, drafted his own bill containing an 8.5 per cent boost for the NIH, from \$15.61 billion to \$16.95 billion.

The bill never made it to the subcommittee hearing room. But it revealed what is likely to be the minimum increase that lawmakers will enact for the NIH, despite extraordinary limits on government spending being faced by the Congress.

In recent years, Congress has not let the figure for the agency fall below that suggested by Porter, an ardent advocate of biomedical research. Last year, he proposed a 9.1 per cent boost. House and Senate negotiators increased that to 14.6 per cent, resulting in a \$2 billion increase for the NIH.

Porter told *Nature* last week that he is hoping for a repeat performance. The 8.5 per cent increase "provides a floor", he said. His ultimate goal is a 15 per cent increase for the NIH, and eventually a doubling over five years — a target supported by the research community and its allies in Congress.

"You have seen powerful evidence that congressional support may be turned into a good increase for NIH," says Mike Stephens, a lobbyist for the Federation of American Societies for Experimental Biology. He adds that the research community should be "buoyed" by Porter's figure.

But Porter cautioned biomedical scientists against assuming that the 15 per cent increase will be achieved. "What you do individually to impact your member of Congress, your senators and the White House is what will make the difference."

Doubling funding over five years would require a 15 per cent boost each year. Such an increase between 1999 and 2000 would require Congress to find an extra \$2.34 billion for the agency this summer.

Porter's Senate counterpart, Arlen Specter (Republican, Pennsylvania), says he is committed to at least a \$2 billion increase for the NIH. But Porter and Specter face formidable budgetary constraints. These make



Porter: suggesting rise of 8.5 per cent.

the agency's fiscal fate uncertain, despite bipartisan support for biomedical research in the House and Senate.

A 1997 budget law imposed caps on non-mandatory government spending, including money for science agencies. So Porter's subcommittee has been allotted \$12 billion less than last year to fund the non-mandatory programmes under its jurisdiction, while Specter's subcommittee faces an \$8 billion shortfall.

Unless Republican congressional leaders agree to break the budget caps, jettisoning the fiscal discipline that they imposed two years ago, it will be virtually impossible to fund such a substantial increase for the NIH. Porter is among Republicans calling for such a move because the government is operating in a surplus for the first time in 30 years.

More significantly, Senator Ted Stevens (Republican, Alaska), chairman of the Senate Appropriations Committee, has indicated his willingness to break the caps.

So many factors remain in play that, until September, when the subcommittees plan to finalize the NIH bill, the rise remains uncertain. Porter says he has been given "strong assurances" by House Republican leaders that they will produce the \$12 billion needed for his bill by September. But they have a razor-thin majority and must juggle a daunting array of considerations, including opposition to breaking the budget caps.

The NIH's prospects remain relatively good, but those of the National Science Foundation (NSF), the space agency NASA and the Environmental Protection Agency are less certain. Their funding depends on a bill that was scheduled to be finalized by the relevant House appropriations subcommittee last Monday (26 July).

The subcommittee is working with a budget allocation \$7 billion below last year's level for non-mandatory programmes. Republicans are trying to use 'emergency' spending and budgetary gimmicks to make up the shortfall while technically remaining within the budget caps.

But Congress is unlikely to match the 9 per cent increase that the NSF received last year. The problem for the NSF is that "everybody likes it but very few love it," says Howard Silver, who chairs the advocacy group the Coalition for National Science Funding.

Meredith Wadman

Michigan to use tobacco money to fund life sciences

[WASHINGTON] The US state of Michigan is to fund a \$50 million-a-year research programme in basic life sciences with money that it will receive from tobacco companies under last year's settlement between the industry and state governments.

The state government hopes that the programme will help to create the critical mass of research needed to nurture a biotechnology industry. Additional impetus will come from major research institutes being built by the University of Michigan and the VanAndel Institute, a private foundation.

Michigan is the first state to dedicate such substantial resources to biomedical research. Scientists hope the idea will be taken up by other states with healthy budget surpluses that are set to receive substantial revenues from tobacco companies.

Governor John Engler said the programme "promises to make Michigan a world-class biotechnology powerhouse". The programme is intended to build on the strengths of the University of Michigan, Wayne State University, Michigan State University and the VanAndel Institute.

Under legislation that was signed by Engler last week, half the money will be directed at collaborations involving at least two institutions, 40 per cent will go to standard investigator grants for basic life science, and the rest will be used help to commercialize discoveries.

The use of tobacco settlement money for biomedical research has been suggested in Washington over the past two years, but Michigan is the first state to put it into practice (see *Nature* 391, 424; 1998).

"The symbolism of using tobacco money for research is very important," said Frank Press, former president of the National Academy of Sciences, speaking at a meeting in Washington at which Engler announced the initiative.

A 14-strong advisory panel, to be appointed by the governor, will draw up details of the programme, which will be administered by Michigan's Economic Development Commission. The first \$50 million will be distributed during the next fiscal year, which starts in October. The law authorizes \$50 million to be spent on the programme each year for 20 years, but actual spending will be subject to annual review by the state government.

Michigan's economy has traditionally been heavily reliant on the motor industry. Although the recent boom in that sector has helped the state, Michigan's political leadership is looking to diversify its economic interests.

Colin Macilwain

Self-navigation put to asteroid fly-by test

[WASHINGTON] A five-minute flight this week by a spacecraft past a tiny, undistinguished asteroid, while not necessarily a great leap for science, could be important for the future of plans by the US space agency NASA to explore the Solar System.

If all goes well, the spacecraft Deep Space 1 (DS1) will today (29 July) pass within 15 kilometres of the centre of the asteroid Braille, far closer than any spacecraft has ever come to a body in the Solar System without landing on it. On-board autonomous navigation, or 'auto-nav', software will direct the 56,000-km-per-hour fly-by with no help from ground controllers, updating the spacecraft's position based on rapidly changing pictures of the target, commanding manoeuvres and deciding when to take photographs.

Since its launch last October, DS1 has shown that it can steer itself during the long cruise phase of a mission using parallax measurements of known asteroids against the star background. The benefit to NASA of this kind of 'hands-off' operation is that the expensive and overbooked Deep Space Network of large radio antennas can be less involved in the routine tracking of spacecraft, and control teams on the ground can be smaller.

The pay-off for science from this week's fly-by will be relatively meagre: around seven black-and-white images with 10-metre resolution, and perhaps 30 more with lower resolution, along with limited infrared imagery and data on the plasma environment surrounding Braille.

But the idea behind the DS1 mission — the first in NASA's New Millennium line of spacecraft — is to demonstrate key technologies such as auto-nav so that more science-orientated missions can incorporate them without being seen as too risky. "We can use it without getting 'dinged' on proposals," says veteran planetary scientist Michael Belton of the National Optical Astronomy Observatories in Tucson, Arizona.

Not everyone in the space science community is convinced of the value of the New Millennium project. It would be better to introduce technological innovation on real, operational science missions, says Stamatios Krimigis, who heads the space science department at Johns Hopkins University's Applied Physics Laboratory (APL), the only institution outside NASA with a major role in managing planetary missions. "Personally, I was never really impressed with demonstrating technology for technology's sake," he says.

With a price tag of \$152 million, DS1 is not much cheaper than APL's \$211-million Near Earth Asteroid Rendezvous (NEAR) science mission, which will orbit the asteroid Eros for months, mapping its surface at 3-metre resolution and returning data on its bulk properties, surface composition and mineralogy across a wide spectral range.

Another Discovery-class mission, the Comet Nucleus Tour (CONTOUR), will fly past three comet nuclei, mapping their surfaces with resolutions down to 4 metres and gathering detailed data on the surrounding



Close call: artist's impression of Deep Space 1's planned encounter with the asteroid Braille.

gas and dust, at a cost of \$154 million. CONTOUR will also reduce its reliance on the Deep Space Network by going into a spin-stabilized 'hibernation' mode during its interplanetary cruise phase.

But even sceptics agree that DS1's successful in-space validation of an ion-drive propulsion engine is an important contribution, as it will make possible certain kinds of spacecraft mission, such as orbiting Jupiter's moon Europa or landing on comet nuclei, that would be impossible or prohibitively expensive with a conventional rocket.

NASA would have been unlikely to approve a scientific mission to an asteroid with DS1's degree of difficulty. "They wouldn't choose a target this small and this poorly known, or go in this close," says Marc Rayman, deputy manager for the mission at the Jet Propulsion Laboratory in Pasadena, California.

Rayman says this week's manoeuvre is "very, very challenging, and it may not work". The approach distance is similar to that of an aircraft flying over the Earth's surface. The spacecraft will not be able to see the mile-wide asteroid until a day before the encounter, and its image will be smaller than one pixel in the camera's view until about an hour before the approach.

By comparison, when the Galileo Jupiter spacecraft flew past two larger asteroids in the early 1990s, scientists on the ground had time to study the approach pictures, plan the photography in advance and radio all the necessary commands to the spacecraft.

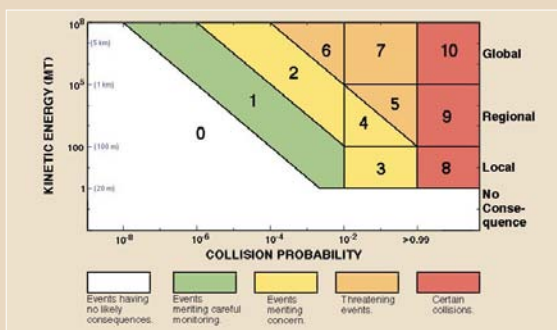
If the autonomous fly-by is successful, DS1 managers have an extended mission up their sleeves, should NASA choose to fund it, which would include encounters with two comets: Wilson-Harrington in January 2001, and Borrelly the following September.

Tony Reichhardt

Scaling the degree of danger from an asteroid

How worried should we be if astronomers were to report finding an asteroid that could collide with Earth 50 years from now? Planetary scientist Richard Binzel of the Massachusetts Institute of Technology has devised the Torino Impact Hazard Scale as a handy tool for scientists, the media and a public confused by a number of recent false alarms (see *Nature* 392, 215; 1998).

Similar to the Richter scale used to rank earthquakes, Binzel's scale plots the energetic 'wallop' packed by an incoming asteroid against the likelihood that it will hit Earth, and assigns numerical values ranging from zero (virtually no chance of damage) to ten



Catastrophe or no? Binzel's Impact Hazard Scale will tell us.

(certain catastrophe) to each event.

Adopted last month at a workshop of the International Astronomical Union in Torino, Italy, the idea was endorsed last week by the full union.

The reviews so far have been good. Carl Pilcher, head of NASA's planetary science

programme, called it "a major advance in our ability to explain the hazard posed by a particular object". Binzel hopes his scale will bring discipline to the reporting of impact hazards. "Scientists haven't done a very good job of communicating to the public," he says.

T.R.

Spanish research to receive funding boost

[BARCELONA] Spain plans to raise spending on research and development (R&D) from its current level of 0.8–0.9 per cent of gross national product (GNP) to 1.2 per cent over the next four years.

This could mean a 10 per cent increase in the government's science budget next year, say officials. The increased funding is promised in a national plan on R&D, outlined to the Senate recently by the minister of education and culture, Mariano Rajoy.

The plan is due to be approved by the cabinet in October. Government officials say the increased effort will be achieved through "slow and sustained increases" in public investment in research, and will be "reviewed annually".

Rajoy told the Senate that the government is keen to encourage more private investment in research, to strengthen the international nature of Spanish science, and to enhance the scientific and technological culture of society.

He said that special emphasis will be put on recruitment and employment issues, including long-term contracts for experienced researchers at research institutes, and support for postdocs to become involved in entrepreneurial activities.

Spain's current level of R&D spending is well below the average of 2.2 per cent of GNP for countries belonging to the Organization for Economic Co-operation and Development. And public spending is still higher than that in the private sector.

The national plan will take account of proposals from Spain's regions, and covers basic and applied research. It has been divided for funding and administrative purposes into one basic research area with no immediate practical applications, and nine scientific and technological areas, including biomedicine, biotechnology, information and communication technologies. There are also 12 sectors considered of special interest to the government, such as defence, energy, space and aeronautics.

Fernando Aldana, director of the Office for Science and Technology, admits that initially more than half of public R&D spending will remain devoted to defence. But he says that the level "will decrease progressively" in favour of civil research. He would like to see total R&D spending in Spain reach 1.5 per cent of GNP in the next decade, with 60 per cent of the investment coming from the private sector.

Although the way the extra money is to be spent is still being discussed, Aldana says that basic research disciplines to be given priority will include astrophysics, particle physics and thermonuclear physics.

He says these relate to areas in which Spain has already made significant contributions to major international projects, including the



Rajoy: wants more private investment.

Gran Telescopio de Canarias, the European Laboratory for Particle Physics, and the Stellerator TJ-II nuclear fusion reactor at the Spanish Centre for Energy, Environment and Technology Research.

Aldana says that private investment will be essential if the government's targets are to be met. Tax relief for corporate R&D programmes will therefore be included in the strategy.

He points out that Spain imports 157 billion pesetas (US\$1 billion) worth of technology annually, but exports only Pts 24 billion worth.

Francisco José Rubia-Vila, director of research of the Community of Madrid, welcomes the planned increase in public spending.

Under rules introduced in December 1998, all laboratories that breed or experiment with animals must register with the Committee for the Purpose of Control and Supervision of Experiments on Animals. The committee is chaired by Maneka Gandhi, the minister for social justice and herself an animal activist.

Although the National Centre for Laboratory Animal Science (NCLAS) has twice applied for registration, this has been refused on the grounds that the centre uses monkeys caught in the wild rather than laboratory bred, and that these are kept in cages that fall below international standards.

The committee's secretary, A. P. Singh, last week ordered NCLAS to release all monkeys back to the wild as a condition of registration. But the centre has refused on the grounds that the monkeys were caught with permission from the forest department, and are being used for research under contract with drug companies.

Without registration, NCLAS chief scientist S. Hariharan says, normal activities at the centre cannot continue. "We have stopped breeding laboratory animals and supplying them to our clients, and have also halted on-going projects," he says.

ing. He says it should allow the "integration of postdocs trained abroad as well as a renewal of infrastructure" and is also confident that private investment will be stimulated.

Francisco Medina-Mena, an electronics researcher at the University of Seville, agrees with the attempt to involve industrial sectors in defining research objectives.

Aldana says one of the novelties of the plan is the "identification of strategic actions involving different sectors". Priorities will be identified for each industrial sector, and scientists will be able to draw on these when putting forward requests for funding.

But others warn that increased funding alone will not improve the functioning of the system. Jesús Martínez-Frías, a geologist at the National Museum of Natural Sciences in Madrid, says there is a need to break down barriers between institutions with responsibility for science. "The functions of each must be defined with clarity and transparency."

He also says that there should be closer coordination with international scientific bodies to avoid excessively 'domestic' policies.

Xavier Bosch

Monkey dispute halts Indian drug tests

[NEW DELHI] A number of drug research projects in India have been halted because of a conflict between a government animal-welfare committee and a Hyderabad centre that supplies animals to research institutes.

Under rules introduced in December 1998, all laboratories that breed or experiment with animals must register with the Committee for the Purpose of Control and Supervision of Experiments on Animals. The committee is chaired by Maneka Gandhi, the minister for social justice and herself an animal activist.

Although the National Centre for Laboratory Animal Science (NCLAS) has twice applied for registration, this has been refused on the grounds that the centre uses monkeys caught in the wild rather than laboratory bred, and that these are kept in cages that fall below international standards.

The committee's secretary, A. P. Singh, last week ordered NCLAS to release all monkeys back to the wild as a condition of registration. But the centre has refused on the grounds that the monkeys were caught with permission from the forest department, and are being used for research under contract with drug companies.

Without registration, NCLAS chief scientist S. Hariharan says, normal activities at the centre cannot continue. "We have stopped breeding laboratory animals and supplying them to our clients, and have also halted on-going projects," he says.



NCLAS is recognized as a centre of excellence by the International Council of Laboratory Animal Sciences and the World Health Organization, and trains scientists from several countries besides India. The centre has more than 30,000 animals, including strains of mice and rats that have been bred over several decades. "If they are lost, it will take another 20 years to establish similar colonies," says Hariharan.

Singh says his committee does not want to harm biomedical research, but he adds: "We have to implement the rules. And I cannot shut my eyes to how the monkeys are kept in NCLAS." But NCLAS scientists say the committee's conditions amount to closing down the facility.

K. S. Jayaraman

US fusion scientists seeking a fresh start

Colin Macilwain

After years of budget cuts and internal division, things are looking up for US fusion research. At a recent meeting, the community began work on a common agenda, and tried to present a united front to funding agencies.

[SNOWMASS, COLORADO] Three hundred US fusion scientists ended a 'consensus' meeting last week aiming to forge a new direction for a field shaken by deep budget cuts and US withdrawal from the International Thermonuclear Experimental Reactor (ITER) programme.

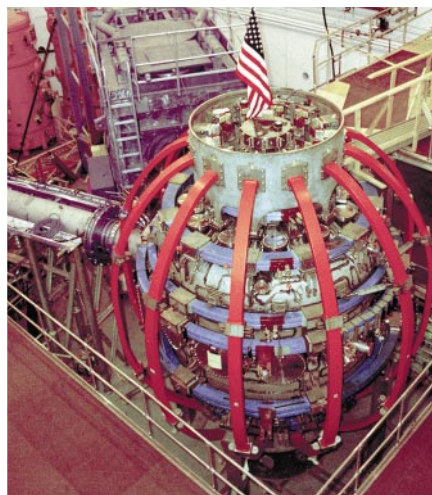
The two-week meeting — an idea borrowed from other branches of physics, but never previously attempted in a field that has traditionally taken its direction from the Department of Energy (DoE) — coincided with word from Washington that the programme's budgetary outlook is improving.

The House Energy and Water Appropriations Subcommittee, which cut the fusion programme's budget in 1995, has marked up an appropriations bill for next year that would expand the fusion programme by more than ten per cent, to \$250 million. It also commended the programme's change of direction.

Although the Senate has proposed a lower budget, fusion advocates are confident that the budget will grow next year. "The very fact that you are having this meeting encouraged the Congressional appropriators' action," Anne Davis, head of the DoE's fusion sciences office, told the meeting in Snowmass, Colorado. "I urge you to keep it up."

Scientists at the meeting agreed general priorities for the various branches of the programme. Perhaps most significantly, they backed a statement that the tokamak is "technically ready for a high-gain burning-plasma experiment" — overwhelmingly rejecting a plea from theorists to qualify this language.

But the meeting's purpose ran deeper than statements of intent. Fusion science in the United States has won a reputation for fractiousness, with scientists and engineers battling over dwindling federal funds (see *Nature* 388, 115–118; 1997). The scientists who organized Snowmass hope that it will mark a rapprochement between these factions.



Great balls of fire: the laser target chamber of the National Ignition Facility, under construction in California, and (top) the spherical torus experiment at the Princeton Plasma Physics Laboratory.

In particular, they hope it will alleviate the tensions that could arise as supporters of inertial fusion energy — which currently attracts only ten per cent of the fusion-energy science budget — seek the backing to exploit the National Ignition Facility (NIF), the \$1.5-billion inertial confinement fusion facility under construction at the Lawrence Livermore National Laboratory in California.

The NIF is being built by the DoE's nuclear weapons programme. But attempts to use it for research into inertial fusion as an energy source will have to be supported from outside the weapons programme, namely by Davis's Office of Fusion Energy Sciences (OFES).

Inertial-fusion researchers would like their share of the OFES budget to grow from \$20 million this year to \$50 million, as they prepare for the NIF to come on line in 2005. They are developing plans for a set of experiments, costing about \$100 million, to explore aspects of inertial fusion outside the NIF's territory — such as the ion beam or laser that would feed power into an inertial-fusion power plant, and the chamber from which heat would be extracted. Scientists at Snowmass were told that these experiments, plus NIF, would provide the technical foundation for an engineering demonstrator of a power plant.

The DoE hopes that the tension between magnetic and inertial approaches to fusion can be reduced by concentrating on areas of common technical interest. At Snowmass, for example, the idea of using 'liquid walls' of molten lithium to absorb neutrons and cool reactor chambers was portrayed as something that could help both approaches. But experienced fusion scientists, such as Miklos Porkolab, director of the C-MOD tokamak at the Massachusetts Institute of Technology (MIT), say that such areas of overlap are rare.

Also discussed at Snowmass was the relationship between the main approaches of the magnetic fusion programme and 'emerging concepts' that would, in effect, respond to the collapse of ITER by taking the programme back to the drawing board.

Non-tokamak approaches to magnetic fusion are being explored with a new vigour. "We deplore the 'down-selecting' that has characterized the programme in the past, and applaud the revival of diversity," says Dan Barnes of the Los Alamos National Laboratory in New Mexico, who led discussions on emerging concepts at Snowmass. But tokamak supporters are not impressed: Porkolab reminded the group that "some of these concepts were discarded 10 or 20 years ago," and warned against "bringing skeletons out of the cupboard".

Technology support, severely neglected in the United States since plans for a fusion materials facility were abandoned in the early 1980s, has also taken on a new lease of life in the programme, and a small materials

PPPL

AP

science programme has been restored. But it remains a minor component of the programme, and the most obvious materials requirement of a fusion energy programme — a neutron source that could simulate the huge neutron fluxes to which a functioning power tokamak would be subjected — remains a distant dream.

The primary goal of the US community has remained constant since before ITER was conceived in 1988: an experiment that would enable physicists to investigate burning plasma.

Few of the scientists at Snowmass dismiss the prospects of Japan, Russia and Europe pressing ahead with a reduced-cost ITER. The Snowmass statement on burning plasma therefore holds open the prospect of the United States being a minor player in ITER, collaborating on an upgrade to JET in the United Kingdom (see *Nature* **389**, 769; 1997), or helping to build Ignitor, the compact design championed by Bruno Coppi of MIT, which is being developed in Italy.

Dale Meade of the Princeton Plasma Physics Laboratory (PPPL) has prepared an outline design for a US burning-plasma experiment, christened the Flexible Ignition Reactor Experiment, or FIRE. Such a step would need major new resources. "No-one wants a burning-plasma experiment to take away from the base programme," says Richard Hawryluk of PPPL, one of the meeting organizers.

And, as John Sheffield of the Oak Ridge National Laboratory, chair of the Fusion Energy Sciences Advisory Council (FESAC), pointed out, no one ruled out some US role in ITER. FESAC met at Snowmass after the consensus meeting to start preparing a key report on the balance of the programme.

Some of the US community feel isolated by the hasty withdrawal from ITER imposed on them by Congress (see *Nature* **394**, 511–512; 1998). "We've erected an iron curtain" by withdrawing so completely, says Charlie Baker of the University of California at San Diego (UCSD), former head of the US ITER home team. He added that the US community is already suffering by not participating in ITER's expert groups, which brought together global expertise in fusion research.

In a sense, the Snowmass meeting was a bid to avoid repeating the mistakes which characterized the entry of the United States into ITER. Parts of the US fusion community never bought into that decision, and their rancour doubtless contributed to Congress's dim view of the international collaboration.

The hope of the Snowmass meeting is that, if the community takes greater control of its destiny, discusses and agrees its priorities and maximizes its scientific output and visibility in a range of fields that could eventually contribute to fusion energy, it will be better positioned to exploit the day when fusion power comes back into fashion.

According to Robert Conn, dean of engineering at UCSD and a member of FESAC, the fortunes of fusion research have always been dominated by externalities — the oil shocks of 1973 and 1979, the thawing of the Cold War in 1988, budget cuts in 1995 and the Test Ban Treaty leading to the construction of NIF. As Meade puts it: "We need to get our sail ready for when the wind changes direction".

External assessment of this strategy has been requested by the DoE. Last month, a sub-panel of the Secretary of Energy's Advisory Board found largely in favour of it, and the National Research Council (NRC) is now

undertaking an additional review. Charles Kennel, director of the Scripps Institution of Oceanography in California and chair of the NRC panel, attended the Snowmass meeting and seemed to like what he saw.

"There's no immediate crisis in the fusion programme," he says. "Clearly when you have a technology development focus you operate like a technology development programme, and when you back off from that, you have a greater opportunity to work with other disciplines. The fact that they captured the idea of the high-energy physicists, to hold a meeting like this without institutional support, is an encouraging sign." □

NIH weighs in behind improved synchrotrons

[WASHINGTON] The National Institutes of Health (NIH) announced last week that it will put \$18 million into "significantly" improving two major synchrotron facilities. The base funding of both facilities comes from the Department of Energy (DoE), and the move reflects the increasing use of synchrotron radiation by life scientists.

Under a memorandum of understanding from the NIH and DoE, the Stanford Synchrotron Radiation Laboratory (SSRL), a division of the Stanford Linear Accelerator Center, will receive \$14 million from the NIH this year, and the National Synchrotron Light Source (NSLS) at the Brookhaven National Laboratory on Long Island, New York, will receive \$4 million.

Synchrotron radiation — bright X-rays produced by electrons circulating in a huge storage ring — gives information on objects at an atomic and molecular level. Biological researchers use it to study protein structure, but drug development is also a key application.

For instance, researchers used DoE-funded synchrotron light sources to develop protease inhibitors for treating HIV infection.

According to Marvin Cassman, director of the NIH's National Institute of General Medical Sciences, the Stanford money is the first instalment of a \$45-million upgrade.



Get the picture: studying protein structure in 3D.

The money will be used to improve the accelerator and instrumentation at NSLS, while the SSRL's electron storage ring will be upgraded to optimize it for biological use. NIH officials point out, however, that the accelerator upgrades at both facilities will benefit all users of the two sources.

In recent years, structural biologists supported by the NIH have become major users of synchrotron sources. This has put pressure on DoE-funded facilities traditionally used by physicists and materials scientists.

But biologists have complained that access to DoE synchrotron sources can take more than six months to arrange (see *Nature* **393**, 3 & **396**, 203; 1998). Last year, a government working group was established to assess how the sources could adapt

to their increasing use by biologists.

The upgrades are one result. "These resources are critical to the development of modern biology, and we feel a responsibility to ensure that they are operated at a level that will benefit the community we serve, and hopefully everybody else," says Cassman, who headed the working group.

NIH director Harold Varmus said that the new money "holds the promise of providing dramatically improved capabilities for determining the structure of important molecules". For instance, the upgrade at the SSRL will enable its five protein-crystallography beamlines each to collect up to five times more data.

Wayne Hendrickson, a structural biologist who is a Howard Hughes Medical Institute investigator at Columbia University, calls the initiative "a very positive move". While recognizing life scientists' growing need for synchrotron radiation, he says, it properly leaves supervision of the expensive facilities in the hands of a single agency, the DoE.

Two newer DoE-funded synchrotron sources will not benefit from the new money. They are the Advanced Photon Source at the DoE's Argonne National Laboratory in Illinois and the Advanced Light Source at the Lawrence Berkeley National Laboratory in California. **Meredith Wadman**

US call for more equitable policy on transplant organs

[WASHINGTON] The US Institute of Medicine last week recommended that scarce transplant livers should be distributed across broad geographic regions, ensuring that the sickest patients receive them, rather than those who happen to live near at hand.

The institute's report supports a policy proposed last year by the Clinton administration, which ordered the United Network for Organ Sharing (UNOS), the group that allocates organs, to distribute them more broadly. The proposal sparked controversy, so Congress requested the report, delaying any change until October.

The report said that "improvements in fairness and effectiveness could be made" to the current system. Donna Shalala, the Secretary of Health and Human Services, said last week that she would put the change into effect quickly. But Congress may give UNOS the power to determine its own policies.

Academies set to give backing to GM crops

[LONDON] Seven academies of science from developed and developing countries have

agreed to develop "an authoritative joint statement" on genetic modification in world agriculture. It is expected to agree that the technology is needed to feed future populations.

The statement is due to be completed by the end of the year by representatives of academies from Brazil, China, India, Mexico, the United Kingdom, the United States and the Third World Academy of Sciences.

Participants at a meeting held this month in London rejected a proposal from the Third World academy to call for an end to the patenting of food crops.

Asteroid on target to give Earth a near miss

[MUNICH] The asteroid 1999 AN 10, predicted to intersect with Earth's orbit in 2027, will not collide with our planet, according to calculations by the German Space Agency. The asteroid will pass 390,000 kilometres from the Earth, equivalent to the average distance between the Earth and the Moon.

1999 AN 10, which has a diameter of about one kilometre, was discovered last January by the US asteroid-search programme LINEAR. But its orbit could not be calculated until it was rediscovered by two German amateur astronomers on archived photographs taken in 1955.

Plea to UK government on cloning research

[LONDON] A complete ban on all cloning research would be "unethical", according to the scientific adviser to Britain's Association of Medical Research Charities. Sir Leslie Turnberg, former president of the Royal College of Physicians, argues in a letter to the *Times* newspaper that cloning research for therapeutic purposes ought to be continued, given its potential to treat diseases.

The letter follows a recent announcement by the UK government that it needs more time to decide whether to allow 'therapeutic' cloning techniques to be used on human embryos, even though they are permitted under current legislation (see *Nature* 400, 4; 1999). Turnberg writes that it would be "shameful" if all forms of cloning research were banned because "society could not trust itself or its scientists to maintain the law".

Unions plan protests at Russian funding gap

[MOSCOW] Russian science has received only two-thirds of the budget it was promised by the government in the first half of this year, says Valery Sobolev, head of the scientific trade unions and social unions co-ordinating committee. Sobolev says his

organization is preparing a series of protests.

The committee says the government promised to borrow US\$5 million from abroad to transfer to scientists, but nothing has been received.

The critics also point out that a revised budget stipulates that only two per cent of state expenditure should be devoted to science, although legislation requires four per cent. Yet military spending is set to increase by 33 per cent.

Japanese bid to end confusion on dioxin limits

[TOKYO] The Japanese parliament last week approved legislation stipulating permissible levels of dioxins — known endocrine disrupters and suspected carcinogens — in air, soil and water. The law, which takes effect early next year, sets standards for the levels of dioxins released from waste incinerators, the main source of emissions, and from factories.

The maximum tolerable daily intake (TDI) of dioxin is set at four picograms per kilogram of body fat. It is hoped that this will help the Environment Agency and the Ministry of Health and Welfare, which currently set different TDI values, to decide on a common standard. The government aims to reduce dioxin emissions by 90 per cent from 1997 levels by 2002 (see *Nature* 398, 362; 1999).

Euro parliament member warns US on trade wars



[STRASBOURG] The newly elected president of the European Parliament's committee on environment, public health and consumer protection has warned the United States not to appeal to 'pure science' alone in trade disputes

over products such as bovine growth hormone and genetically modified crops.

British Conservative Caroline Jackson said last week that there is "a big difference between Europe and the United States on this type of thing". She said her committee may decide to address the extent to which the World Trade Organization is required to use judgements on scientific validity as a basis of adjudicating on trade disputes.

Jackson is an active member of the European People's Party, a grouping of centre-right candidates which won the largest number of seats in last month's parliamentary elections. During the allocation of committee responsibilities last week, Carlos Westendorp Y Cabeza, a Spanish socialist, was elected chair of the

parliamentary committee that oversees the European Commission's Framework research programmes.

Nobel laureate Mullis hangs up his surfboard

[SAN DIEGO] Nobel laureate Kary Mullis, the inventor of the polymerase chain reaction (PCR) used to amplify DNA, has joined a medical diagnostic firm in Irvine, California. Since winning the prize in 1993, Mullis has led a peripatetic life, consulting for biotech firms, writing, surfing and skating.

Mullis, now director of molecular biology at Burstein Laboratories, is overseeing the development of laser-disc technology for medical analysis. "Everyone has to take a job sometime," says Mullis, who had been living off the Nobel award. When he devised PCR, Mullis was a researcher at Cetus Corp., which reaped virtually all the economic benefit.

correction: French genome research

Our article on funding for French genome research stated incorrectly that money allocated to a public/private consortium would be used to develop a national network of 'genopôles' (*Nature* 400, 199; 1999). In fact, these clusters of genome-related labs and companies will be supported through a separate initiative.

Precautionary principle stifles discovery

Sir — The so-called 'precautionary principle' (PP) has gained currency in discussions about environmental protection and genetic manipulation, but it should be treated with caution.

The principle has been endorsed in international treaties, including the consolidated version of the treaty establishing the European Union. In many of these documents the PP has not been explicitly defined, but the Wingspread conference attempted to define it¹. We believe the following definition would be accepted by most proponents:

"When an activity raises threats of serious or irreversible harm to human health or the environment, precautionary measures that prevent the possibility of harm (for example, moratorium, prohibition) shall be taken even if the causal link between the activity and the possible harm has not been proven or the causal link is weak and the harm is unlikely to occur."

In our view, there are problems with the

PP as so defined. The PP tells us to balance evidence in a specific way. The weight given to evidence is ordinarily thought to be a function of its epistemic warrant (the degree to which we have reasons for believing the evidence). The PP instructs us to change this normal balancing by giving evidence pointing in one direction more importance than evidence pointing in the other direction, even in cases where the evidence has the same epistemic warrant. Such discounting will distort our beliefs about the world, and will lead us to hold false beliefs. The PP cannot therefore be a valid principle for evaluating evidence.

As a principle of rational choice, the PP will leave us paralysed. In the case of genetically modified (GM) plants, for example, the greatest uncertainty about their possible harmfulness existed before anybody had yet produced one. The PP would have instructed us not to proceed any further, and the data to show whether there are real risks would never have been

produced. The same is true for every subsequent step in the process of introducing GM plants. The PP will tell us not to proceed, because there is some threat of harm that cannot be conclusively ruled out, based on evidence from the preceding step. The PP will block the development of any technology if there is the slightest theoretical possibility of harm. So it cannot be a valid rule for rational decisions.

This fatal weakness of the PP illustrates a common problem in attempting to convert moral choices into legislation. The temptation is great to try to find one absolute and easily applicable principle, but such a principle will often be simplistic and will, when applied, lead to unjustifiable conclusions. Many moral choices are complex, and in making political decisions we should not lose sight of this complexity.

Søren Holm, John Harris

Institute of Medicine, Law and Bioethics,

University of Manchester, Manchester M13 9PL, UK

1. <http://www.wajones.org/wingscons.html>

Funding agencies must use their muscle

Sir — Your decision to let authors cite their contributions to papers is a welcome change in policy (*Nature* 399, 393; 1999). Statements clearly allocating credit and responsibility for the research done can only help to promote the health of science.

But it is likely that many authors will need persuasion before they embrace the idea of citing contributions. Such persuasion is unlikely to come from the journals themselves. Although some (notably *The Lancet*) courageously require the contributions of each author to be cited in papers, most do not — presumably for fear of alienating their clientele. It is probably too much to expect journals to threaten their livelihoods by imposing a rule that is unpalatable to many (particularly to senior) scientists.

But such persuasion might legitimately come from the agencies that fund research. If the US National Institutes of Health and the UK Wellcome Trust, for example, were to require every grant recipient to cite each author's contribution to any paper resulting from the research funded, this would promote the adoption of the new authorship policy. Because the major funding agencies need fear no reprisal from researchers who object to the policy, and because they have a vested interest in maintaining the health of the scientific

enterprise, they stand in a unique position to bring change in this important area.

Benjamin White

Department of Pharmacology, Yale University School of Medicine, New Haven, Connecticut 06520, USA

Substance versus style in scientific papers

Sir — Antonio J. Herrera complains about referees' criticisms of the English of papers by non-anglophones (*Nature* 397, 467; 1999). Scientists of any nationality can produce verbose and dull writing in their mother tongue, although it is naturally harder to write in a foreign language.

But manuscripts with outstanding scientific content are never rejected because of the prose. If you think that the English in your paper has been overcriticized, or that rejection is based on judgements of style, you should inform the editor.

Rather than dreaming up an Institute for Correct English Style, I believe all scientists should be taught to write while at university. Students should learn the attributes of good writing, grammar and composition, and the logical development of scientific argument. Once these skills have been ingrained in one's native language, faults in English will be perceived and corrected.

Marta Pulido

Institut Municipal d'Investigació Mèdica,

Doctor Aiguader 80, E-08003 Barcelona, Spain

Flood warnings

Sir — Proffering the tantalizing nomenclature of zettabytes and yottabytes as though they were exotic new species, you warn of the tidal wave of data that threatens to engulf science¹.

We read that, while the physics and remote-sensing communities are getting ready to ride the wave, "many biologists are still in denial". This may be so, but visionary biologists support the inspiring concept of a Global Biodiversity Information Facility (GBIF)² that has been developed by the Megascience Forum of the Organization for Economic Cooperation and Development. This distributed facility has been proposed in response to the challenges of the vast domain of biodiversity information. It aims to provide orderly and structured access to our knowledge of the millions of species, and is vital if we are to keep pace with, for example, our burgeoning knowledge of genomes.

Every deluge needs a Noah's Ark. How successfully biology survives the flood will depend upon recognizing that systematic biology underpins our ability to understand the living world³, whether it is represented by genome sequences or ecosystems. We must build and float the GBIF soon.

Stephen Blackmore

Department of Botany, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

1. Reichhardt, T. *Nature* 399, 517–520 (1999).

2. <http://www.york.biosis.org/gbif/index.htm>

3. http://www.nhm.ac.uk/hosted_sites/ukfs/web_of_life/index.htm

A manual for planetary management

It is time for environmental scientists and policy-makers to speak the same language, and to target the achievable, not simply the desirable. A framework is emerging from the International Geosphere–Biosphere Programme.

Philip Newton

When it comes to global environmental change, scientists and policy-makers rarely interact effectively. The reasons are not difficult to understand. Policy-makers prefer to make decisions, especially expensive ones, based on a degree of certainty, yet such is the elusive complexity of the Earth system that scientists are often able to do little more than outline a broad range of possibilities.

Witness the situation regarding future anthropogenic emissions of greenhouse gases: scientific advice on the likely impacts largely amounts to a wide range of possible consequences along with a warning that surprises are likely. It is hardly surprising that the largely politically brokered Kyoto Protocol amounts to little more than a tiny step towards limiting climate change¹.

So how can scientists increase the chances of effective management of global change? Clearly there is a need to improve understanding of how the Earth system functions, but how can this be focused so as to optimize the role of policy-makers? Such questions were a strong motivation for the second congress of the International Geosphere–Biosphere Programme (IGBP), held in May².

The business of the IGBP is global-change research, which demands large-scale interdisciplinary actions. The IGBP's strategy over the past decade has been to coordinate activities within and between independently established international 'core' research programmes covering atmospheric chemistry, terrestrial and marine ecosystems, land-use change, hydrology, land–ocean interactions, ocean biogeochemistry and past climate change². This strategy has had successes. For example, as a result of such coordination it is now generally accepted (and widely touted) that the anticipated carbon sequestration by land vegetation in the coming century or so is likely to hit a ceiling. The identified limiting factors include the diminishing response of photosynthesis to an increasing burden of atmospheric CO₂, a temperature-driven increase in respiration of plant carbon in soils, and, in systems where vegetation replacement is induced, the slow rate of carbon gain during growth relative to that during vegetation loss.

But, although many research communities have mixed and broadened constructively, such as between hydrological and land-use-change communities, there remains an urgent need to stretch further. For example, are physical and social scientists, or even terrestrial and marine ecologists, yet speaking the same language? And, if understanding

the past is the key to the future, to quote the palaeoscientist's maxim, why has the integration of palaeoscience with other IGBP programmes been rather limited?

To tackle such issues, there is a need to map out a new strategy for global-change research. First, if the prerequisite of an improved understanding of the Earth system is to be attained, then research must become yet more interdisciplinary. Second, the focus needs to be on achievable and policy-relevant aims over the next decade — addressing questions we feel we can realistically answer, not those we would like to just out of curiosity. Third, a framework for integration of the observational, experimental and modelling activities is needed from the outset, to guide strategy. Leaving the integration until the end has done a fine job of showing where efforts should have been focused, when it is too late to fill the gaps. And, as Pam Matson of Stanford University puts it, funding across disciplines should be "matrix money, not glue money".

These are laudable aims, but what do they imply on a practical level? As a first step, most IGBP programmes have embarked on a phase of synthesis. This will elucidate the way forward — what do we know, what do we think we know, and what do we need to know, to achieve goals relevant to policy? Part of the 'need to know' challenge must include how to deal with up- and down-scaling in space and time. This process is intended to spawn the programmes required to attain such goals.

It will be very hard to establish an integration framework, particularly on a global scale, as present capabilities for modelling the Earth system are rather limited. A dual approach is planned. On the one hand, the relatively conventional approach of improving coupled atmosphere–ocean–land–ice models will be pursued. Ingenuity aside, the computational demands are extreme, as is borne out by the Earth System Simulator — 640 linked supercomputers providing 40 teraflops and a cooling system from hell under one roof — to be built in Japan by 2003.

The second course is the development of Earth-system analysis, such as proposed by John Schellnhuber³, as a hypothesis-testing framework. Such analyses aim to include the social dimension, ultimately aspiring to, as Schellnhuber puts it, a mathematical analysis of sustainable development, leading to the

notion of a 'manual' for planetary management. A policy-relevant aim might be to develop such models to identify the likely pathways of human activities that could lead to intolerable risks to the global environment, so as to chart probable safe courses.

Integration on local to regional scales for problem-solving studies is somewhat less daunting. But if a region undergoing environmental change is to be investigated, the strategy from the outset should involve scientists across all the relevant disciplines, from the physical to the social, from ecology to economics. Such an approach is exemplified by the Large-scale Biosphere–Atmosphere experiment in Amazonia⁴, which embraces many IGBP core programmes.

A framework for future studies is emerging. The emphasis would be on environmental change of relevance to society, at global and regional (but not national, to lessen political issues) scales, largely focused on the next 10–20 years, and linking the 'hard' sciences to the social. Centred on the

Earth system as a whole, cross-cutting activities would target themes such as the carbon and water cycles, or ecosystem services. New activities at existing disciplinary interfaces would be encouraged. Plans are in place to improve crucially needed links to other international organizations such as the World Climate

Research Programme, the International Human Dimensions Programme and the Food and Agriculture Organization.

This is an ambitious enterprise, and whether it stands or falls rests on not only the skill, but the enthusiasm, of the scientists involved. Crossing disciplinary barriers requires considerable effort, both intellectually and in procuring funds from organizations that are largely structured to operate by discipline. The IGBP has no crock of gold to back up its fine words. Moreover, much of the measurement and observational work will have less allure than curiosity-driven research. The motivational role of the IGBP should not be underestimated. □

Philip Newton is a senior editor at Nature and a visiting research fellow at the School of Biological and Molecular Sciences, Oxford Brookes University, Oxford OX2 0BP, UK.

1. Parry, M. et al. *Nature* **395**, 741 (1998).

2. <http://www.igbp.kva.se>

3. Schellnhuber, H.-J. & Wenzel, V. (eds) *Earth System Analysis: Integrating Science for Sustainability* (Springer, Berlin, 1998).

4. <http://www.cptec.inpe.br/lba>



HERBERT GIRAUDET/STILL PICTURES

Rebuilding the road to cancer

Jonathan B. Weitzman and Moshe Yaniv

After more than 15 years of trying, researchers have managed to convert normal human cells into tumour cells in a culture dish. This achievement should help identify new players in tumour formation, and develop treatments that target them.

We know that the pathway to tumour formation is a multi-step journey involving the accumulation of genetic alterations along the way¹. We also know much about the cellular hurdles on this road, and the mutations that allow tumour cells to overcome them. But do we know enough about cancer to allow us to convert normal human cells into fully fledged tumour cells in a culture dish? We do now. Reporting on page 464 of this issue², researchers in Robert Weinberg's laboratory have combined all that we have learned to effectively turn human cells into tumour cells in the laboratory. They set out to determine the minimum number of defined genetic events required for tumour formation, and have brought an end to an endeavour that began more than 15 years ago.

Normal cells have a number of intrinsic mechanisms, involving molecular 'gatekeepers', to protect them from going out of control and forming tumours. All cells have a limited ability to replicate themselves, and, when grown in culture, will eventually reach a non-dividing state called senescence³. But genetic alterations allow some cells to escape senescence. These cells pass through a 'crisis' phase, which is usually accompanied by extensive cell death, and they become immortalized — they continue to proliferate indefinitely. They may then acquire the characteristics of cells that are fully cancerous — or 'transformed' — through additional mutations in tumour-suppressor genes or oncogenes (genes that cause cancer when mutated). Unlike normal cells, transformed cells can survive without growth factors, and they do not need to be anchored to a solid support.

The concept that genetic events cooperate to achieve transformation was proposed over a decade ago. Cuzin, Weinberg and others^{4–6} showed that to transform normal rodent cells to a cancerous state, delivery of two oncogenes was required. In contrast, immortalized rodent cell lines could be transformed by a single oncogene. But when researchers tried similar experiments with human cells they consistently failed to obtain transformed cells, highlighting a mechanistic difference between rodents and humans in terms of cellular control.

It has recently become clear that these dif-

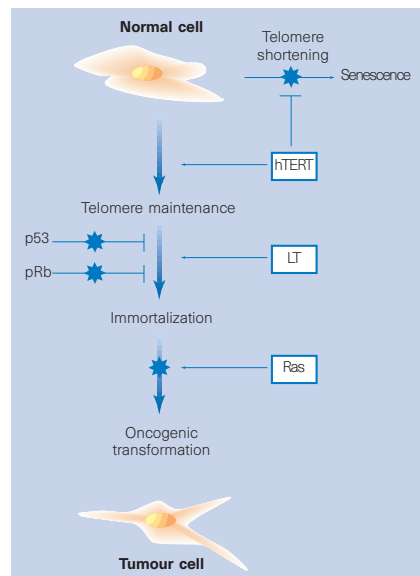


Figure 1 Genetic events required to convert a normal human cell into a tumour cell. Weinberg and colleagues² show that alterations in at least four pathways are needed. Tumour formation can be mimicked in the laboratory by delivering the catalytic hTERT subunit of telomerase (which maintains telomere length), combined with SV40 large-T antigen, LT (which inactivates both the p53 and retinoblastoma (pRb) pathways), and an activated *ras* oncogene (which induces transformation to a cancerous state, allowing cells to grow indefinitely in the absence of growth factors).

ferences are due to telomere biology. Telomeres are specialized structures that define the ends of chromosomes. With each round of cell division, the DNA at each telomere gets progressively shorter, leading, in turn, to shortening of the whole telomeric structure. Several observations suggested that this telomeric erosion could impose the limits on cellular lifespan. Normal human tissues have low levels of telomerase, an enzyme that is responsible for maintaining telomere length. In contrast, many tumour cells and immortalized cell lines have upregulated telomerase activity, suggesting a mechanism by which cells may escape normal ageing controls⁷. Mouse cells have telomerase activity and also longer telomeres — this could explain the differences between human and rodent cells

in experimental transformation assays.

Weinberg and colleagues² predicted that if telomere erosion were prevented, allowing cells to become immortalized, this could cooperate with additional oncogenes and allow direct progression to cancer. The players that the authors used in this story are well known to cancer biologists, but it was the combination of these elements that ensured the successful tumour formation. The authors began with normal human epithelial cells and fibroblast cells, to ensure that their observations were not confined to one particular type of cell. To overcome telomere shortening and bypass senescence and crisis, they artificially expressed the telomerase enzyme's catalytic subunit (known as hTERT). This approach has already been shown to immortalize human cells^{8–11}. Weinberg and colleagues then transformed these immortalized cells by adding two familiar oncogenes. The first is an activated *ras* gene (H-*ras*V12), which is mutated in many human tumours. The second is a viral oncoprotein, the simian virus 40 large-T antigen. This protein is involved in transforming normal host cells into tumour cells by inhibiting two central growth-control factors — the tumour-suppressor proteins p53 and retinoblastoma.

When the authors introduced the *ras* oncogene alone, the human cells became senescent (as previously reported for mouse cells). This senescence could not be overcome by combining *ras* with introduction of hTERT. But cells expressing large-T and hTERT became immortalized, and when these proteins were combined with *ras* expression the authors saw slightly increased growth of the cells. Weinberg and colleagues then tested the ability of these different cells to form colonies in soft agar — a simple, *in vitro* test for whether cells need to be anchored to a solid surface for growth (normal fibroblasts and epithelial cells do, whereas transformed cells do not). The authors also examined the ability of the cells to form tumours *in vivo* by injecting them into 'nude' mice, which have a defective immune system. The results clearly showed that only the combination of all three genetic elements was sufficient to cause efficient tumour growth. The authors conclude that mutations causing telomerase to become active can cooperate with other mutations in oncogenes to induce transformation.

What is the minimum number of 'genetic hits' required for tumour formation? At least four distinct pathways are involved (Fig. 1), as the large-T protein touches both the retinoblastoma and p53 pathways. This is one more pathway than is required for transformation of rodent cells. However, the issue remains controversial. Earlier this year, Morales *et al.*¹¹ reported that they could not transform human fibroblasts using hTERT and *ras* combined with the human papillo-

mavirus-16 E6/E7 viral oncoproteins (which also inactivate the retinoblastoma protein and p53). It is possible that large-T perturbs additional cellular targets. Or perhaps the order in which the mutations occur is important. Weinberg and colleagues began their protocol with the large-T protein, whereas Morales *et al.* first created cells that expressed hTERT. Inactivation of p53/retinoblastoma functions increases genomic instability (leading to mutations), whereas switching on hTERT does not.

These issues need to be investigated — by changing the order of gene delivery, or by using large-T variants with defined characteristics — to clarify just how many hits (four, five or more?) are required. Nonetheless, this landmark paper should spur other labs to develop their own *in vitro* tumour formation kits, allowing them to dissect the roles of individual oncoproteins in transformation. Finally, transforming cells in dishes is all very well but, *in vivo*, cancer relies on the tumour's ability to evade the immune

system, to attract its own blood vessels and to spread around the body. So although attempts to create human tumour cells have now been brought to an end (of the chromosome, as it turns out), the challenge to integrate knowledge from *in vitro* transformation models with an understanding of complex, multi-step tumour formation *in vivo* is far from over. □

Jonathan B. Weitzman and Moshe Yaniv are in the Unité des Virus Oncogènes, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France.

e-mails: jonnyw@pasteur.fr

yaniv@pasteur.fr

1. Kinzler, K. W. & Vogelstein, B. *Cell* **87**, 159–170 (1996).
2. Hahn, W. C. *et al.* *Nature* **400**, 464–468 (1999).
3. Hayflick, L. & Moorhead, P. S. *Exp. Cell Res.* **37**, 614–636 (1965).
4. Land, H., Parada, L. F. & Weinberg, R. A. *Nature* **304**, 596–602 (1983).
5. Rassoulzadegan, M. *et al.* *Nature* **300**, 713–718 (1982).
6. Ruley, H. E. *Nature* **304**, 602–606 (1983).
7. Meyerson, M. *et al.* *Cell* **90**, 785–795 (1997).
8. Kiyono, T. *et al.* *Nature* **396**, 84–88 (1998).
9. Bodnar, A. G. *et al.* *Science* **279**, 349–352 (1998).
10. Counter, C. M. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 14723–14728 (1998).
11. Morales, C. P. *et al.* *Nature Genet.* **21**, 115–118 (1999).

Galaxy formation

Mysteries of the Milky Way

Gerry Gilmore

The outer reaches of the Milky Way galaxy contain the oldest known stars. Nonetheless, the birthplace of these stars, and the time when they became members of the Milky Way, remain the subject of controversy. The simplest view is that the stars formed more or less where they are now, and are the long-lived fossil remnants of the earliest stage of galaxy formation. On the other hand, the most successful models of structure formation in the Universe, those involving 'cold dark matter' (CDM), have stars forming in small dwarf galaxies, which are only much later accreted into larger galaxies — old stars in a new home. Yet, although CDM models are rather successful on large scales, they are poor at describing individual galaxies. The models should work best in the outer parts of galaxies, where one expects the fragile, dwarf galaxies predicted by CDM models to have left behind some of their stars as they are tidally devoured by the Milky Way over many billions of years. A study by Alex Stephens¹ in the *Astronomical Journal* looks at the relative abundances of chemical elements in stars that are currently near the Sun, but which orbit far into the outer Galaxy. This work has greatly improved the available evidence, but at the same time created more problems for CDM models.

High-mass stars live short lives, during which they create new chemical elements. Stars less massive than the Sun live for longer, and retain in their atmospheres the original distribution of chemical elements in the gas from which they formed. Modern spectro-

scopic analyses of low-mass stars with a range of ages can therefore tell us the evolutionary history of these chemical elements. Stars outside the disk of the Milky Way, but within an area known as the Galactic halo, are some of the oldest, and are thought to be part of the original Milky Way, before much of it collapsed to a disk. Others argue that these halo stars may have been captured from satellite galaxies, but the evidence is inconclusive.

The task of working out the rate at which halo stars formed in the early Universe — which cannot be detected directly — is eased

by a fortunate piece of physics. The so-called α -elements (which are created by adding α -particles to the nucleus), comprising oxygen, magnesium, silicon, titanium and calcium, are created exclusively in high-mass, very short-lived stars, and are ejected by a type II supernova explosion. Some heavier elements, most importantly iron, are created in type I supernovae, whose poorly understood progenitors are believed to have a pre-explosion lifetime of typically one billion years (1 Gyr). So, comparing the relative abundances of the α -elements and iron provides one with a clock, with a resolution somewhat better than 1 Gyr. It is this clock that Stephens¹ has applied to halo stars visiting the solar neighbourhood from the far outer Galaxy.

Previous studies of halo stars near the Sun, mainly on orbits from the inner halo, show them to have a ratio of α -elements to iron that is a factor of about two higher than that of the Sun. This shows that all these stars formed within about 1 Gyr of the beginning of star formation, and it is consistent with simple, rapid, *in situ* formation of the inner halo. But is it also consistent with CDM models of late accretion of small, old galaxies? The snag here is that no one knows how to incorporate star-formation theory into galaxy-formation models in a realistic way that reproduces observations. One way to test the predictions of CDM models might be to assume that the small satellite galaxies around the Milky Way are remnants of the original CDM structures from which the Milky Way later formed.

This approach gained impetus following the discovery of the Sagittarius dwarf (Fig. 1), a small galaxy deep inside the halo of the Milky Way, which behaves exactly like a CDM fragment: it shows every sign of being tidally disrupted and merging into the Milky Way². By comparing the observed distribution of stellar ages and predicted chemical abundances in such satellite galaxies with

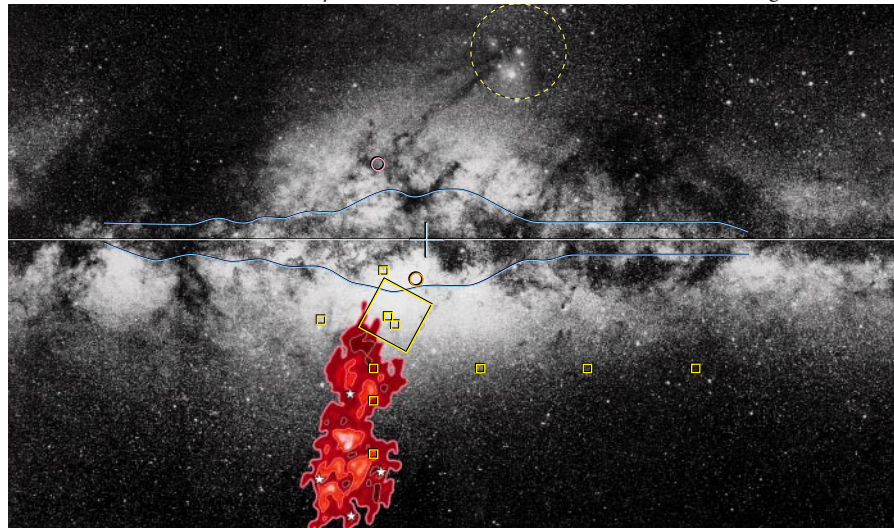


Figure 1 Optical image of the central Milky Way galaxy. The irregular shape outlines the Sagittarius dwarf galaxy, which is inside the halo of, and probably being assimilated into, the Milky Way. Such mergers are predicted to be common, but on recent evidence (notably, chemical analysis of outer-halo stars by Stephens¹) seem to be quite rare. Circles indicate areas studied for comparison with Stephens's halo stars.

RICHARD SWORD, INSTITUTE OF ASTRONOMY

those for Galactic-halo stars, it is possible to test for a common origin. The inner-halo stars fail this comparison spectacularly: the satellite galaxies are systematically much younger, so the inner halo does not seem to have been formed by late aggregation of many small satellite galaxies.

That said, the real test of late-accretion models involves the stars of the outer halo, because it is out there that fluffy little galaxies will be shredded by Galactic tides, during the many mergers predicted by CDM theory. Stephens¹ provides the first high-quality results for a useful sample of stars with indisputable outer-halo orbits. Surprisingly, these stars are indistinguishable, in both their relative and absolute distributions of iron and α -elements, from the stars of the inner halo. The similarity in the absolute distributions is intriguing, as it is at odds with all simple models of galaxy formation in which a gaseous galaxy collapses slowly, forming stars in dense, short-lived gas clouds during the collapse. Such models predict that the outermost stars form earlier, on average, and so should have lower abundances of all the elements, because they formed before there had been time for many element-creating supernovae. The similarity of the relative abundances, as outlined above, is hard to reconcile with late-accretion models.

The most straightforward interpretation

of all the relevant data on halo stars, including the new results from Stephens, is that the first significant star formation began in an extended but somewhat centrally condensed region, in a large number of gas clouds. Moreover, the beginning of star formation must have been synchronized within a period of 1 Gyr or so, across a region the size of the present Milky Way halo. This is not impossible, considering that a typical halo gas cloud will cross the entire halo diameter (roughly 50,000 light years) twice in that time, allowing for at least one star-forming collision. These clouds must have been sufficiently massive to survive one or more generations of star formation, supernovae and chemical-element enrichment. Whether it is more appropriate to call this inhomogeneous collapse, rapid early merger activity, or some other popular model, is perhaps more relevant to linguistics than to astrophysics. Perhaps, over the rest of the history of the Universe, a few dwarf galaxies — in much smaller numbers than predicted by current CDM models — strayed close to and were captured by the Milky Way, a process that continues today. □

Gerry Gilmore is at the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK.
e-mail: gil@ast.cam.ac.uk

1. Stephens, A. *Astron. J.* 117, 1771–1791 (1999).

2. Ibata, R., Gilmore, G. & Irwin, M. *Nature* 370, 194 (1994).

Microbiology

What's eating the free lunch?

Gary J. Olsen

If there is a free lunch to be had, someone (or something) will eat it. Inspired by faith in this principle, researchers have analysed the chemical energetics of many combinations of naturally occurring oxidants and reductants. Then, for each combination that is energetically favourable, they have sought an organism that exploits the reaction for its growth. Such searches have not always been successful. For example, over 20 years ago it was asserted that an organism that can use nitrite to oxidize ammonium anaerobically, in the so-called anammox reaction $\text{NH}_4^+ + \text{NO}_2^- \rightarrow \text{N}_2 + 2\text{H}_2\text{O}$, was 'missing' from nature¹.

However, the inability to cultivate a particular type of organism is no proof that it does not exist. Indeed, in this case, perseverance has been rewarded — on page 446 of this issue², Strous *et al.* describe the cultivation (in an enriched, but still mixed, culture) and partial characterization of a long-sought anammox organism. The results are of both scientific and practical value. Scientifically, there is the general question as to why nature is so good at producing organisms to exploit almost every free lunch, and, more specifically, the matter of how the organism deals

with some highly reactive intermediates in the reaction. On the practical side, the organism might be used to lower the fixed nitrogen content of waste waters, a major concern in freshwater environments.

So what is this organism? Strous and colleagues' comparisons of molecular sequence data (ribosomal RNA genes) indicate that it is a distant relative of the planctomycetes, a group that includes the obscure genera *Planctomyces*, *Pirellula*, *Gemmata* and *Isosphaera* (see their Fig. 3 on page 448). Although planctomycetes belong to the Bacteria³, they have some unusual features: cell division by budding (reminiscent of some yeasts), complex internal membrane structures (reminiscent of eukaryotes) and cell walls that lack peptidoglycan (reminiscent of Archaea)⁴. Even though the anammox organism is only distantly related to previously cultivated planctomycetes, it also divides by budding and has internal membranes. (The cell-wall composition is not reported.) Strous *et al.* speculate that the membrane structure might protect the cell from the reaction intermediates hydrazine (a good rocket fuel) and hydroxylamine (a



100 YEARS AGO

An ingenious machine for printing in colours, invented by Mr. Ivan Orloff, chief engineer and manager of the Russian Government printing works at St. Petersburg, has just been set up in London, and a company has been formed to develop the use of the machine for supplying coloured illustrations for periodicals and books. In colour printing by the ordinary method the successive colours are applied one at a time as the preceding one becomes dry. By means of the Orloff machine the whole of the colours required in a picture are printed at a single turn of the cylinder. If the picture has to be printed in, say, four colours, four separate blocks are arranged around the curved surface of the cylinder. As each block passes a particular point, the roller carrying the colour required by the block is made to fall upon it by a system of cams. Each block thus receives the coloured ink intended for it in the course of a revolution of the cylinder. ... The results are very effective, and the "register" is perfect, no matter how many colours are used. The machine appears to mark a distinct development of methods of printing in colours.

From *Nature* 27 July 1899.

50 YEARS AGO

Luminous night clouds were observed in the northern sky in Scotland on the nights of July 9–10 and July 10–11, the extensive display showing characteristic wave structure in the early morning of July 11 being quite magnificent. Brilliantly blue-white in colour, the night clouds appeared about half an hour after sunset in a sky almost clear of normal cloud, were most brilliant during the hour after midnight and faded in the growing light before sunrise. ... These rare clouds, which are observed only in summer and in middle latitudes, were first studied by Jesse and later by Störmer using his auroral network. ... Occurring at heights of about 80 km., they provide valuable information concerning air movement at that level. On the evidence of their colour, it is generally believed that they consist of fine dust of cosmic or volcanic origin. It is significant that the present manifestation followed by a few days after an eruption in the region of the Canary Islands.

From *Nature* 30 July 1949.

good mutagen), but this seems unlikely as both of these chemicals readily cross membranes. For now, we neither know the function of the membranes, nor the mechanisms of cellular protection.

In contrast to these unifying structural features, the metabolism of the anammox organism differs dramatically from that of the other planctomycetes^{2,4}. The organism is a lithotroph, deriving its energy from inorganic compounds (ammonium and nitrite); other planctomycetes derive their energy from the oxidation of organic compounds. It is also an autotroph, able to fix carbon dioxide for its cellular carbon; other planctomycetes obtain their carbon by assimilating organic molecules. And it is anaerobic, respiring nitrite; the other planctomycetes are aerobic, respiring oxygen (although *Pirellula* species can also respire nitrate). These differences emphasize the tremendous metabolic diversity in related microorganisms.

Physical and biological reactions provide many potential sources of energy. The question is whether organisms can exploit those sources before they are dissipated by physical processes. To use an energy source, an organism must first be able to catalyse all or part of the path to equilibrium (an ability to eat the food); second, it must have a method to couple the equilibration with some biologically useful process or reaction (an ability to digest the food). Can evolution produce every useful metabolism? Are there limits on the inventiveness of life, or is everything possible through evolution?

I suspect that there are limits, but evolution is opportunistic and we easily underestimate the opportunities. If a resource is available to a large population (giga-Petri-dish-size experiments) for a long period of time (tens or hundreds of mega-graduate-student tenures), much can happen. Natural selection is powerful, and we do not see its countless failures. In addition, evolutionary resources transcend organismal and species boundaries; physiological capabilities can be drawn from any of thousands of microbial species by dividing steps of a reaction among two or more organisms in the community or by passing genes among organisms. In this regard, Strous *et al.* find that the anammox organism must be cultivated with other cells. The contribution of these neighbouring organisms is unknown, but it is not part of the anammox reaction as such.

We can push the limits of the free-lunch idea. An observation reported last year is that, in the hydrothermal-vent environment of organisms such as *Methanococcus jannaschii*, the synthesis of the more hydrophobic amino acids is energetically favourable⁵. For example, the synthesis of methionine from ambient H₂, NH₃, CO₂ and H₂S has a net free-energy of about -175 kJ mol⁻¹. That is, unlike ourselves, or *Escherichia coli*, *M. jannaschii* could in principle synthesize

methionine and several other amino acids with no input of cellular energy; in fact it could make an energetic profit. Thinking about it naively, why shouldn't these organisms, or others in comparably highly reduced environments (mostly due to high H₂), have discovered that they can have a free lunch by making certain amino acids (E. L. Shock, personal communication)? At present, we have no answers. But given the existence of the anammox organism, and of others that exploit 'unusual' resources, it seems

that we need to broaden our thinking, lest we overlook some very exciting biology. □

Gary J. Olsen is in the Department of Microbiology, University of Illinois, Urbana, Illinois 61801, USA. e-mail: gary@phylo.life.uiuc.edu

1. Broda, E. Z. *Allgem. Mikrobiol.* **17**, 491–493 (1977).
2. Strous, M. *et al. Nature* **400**, 446–449 (1999).
3. Woese, C. R., Kandler, O. & Wheelis, M. L. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
4. Staley, J. T., Fuerst, J. A., Giovannoni, S. & Schlesner, H. in *The Prokaryotes*, 2nd edn (eds Balows, A. *et al.*) 3710–3731 (Springer, New York, 1992).
5. Amend, J. P. & Shock, E. L. *Science* **281**, 1659–1662 (1998).

Condensed-matter science

Magnetic phases to order

Neil Mathur

Condensed-matter physics met chemistry head-on at a workshop in Russia earlier this month*, where an impressive array of experimental and theoretical evidence showed that phase separations occur over a variety of length scales, and that the various states of matter that arise can be both physically and chemically tuned.

Complex magnetic oxides, such as manganites of the form (La,Pr,Ca)MnO₃, can be ordered or disordered over different length scales. A fascinating range of spin, charge and orbital states can thus exist to different spatial extents. Various combinations of these states might order, disorder or melt to become dynamic. Chemists might look at

the short-range correlations between electronic spins, charges and orbitals (bonds). These can collectively produce long-range effects leading, for example, to new structural symmetries (Fig. 1a). Long-range correlations can also lead to metallic behaviour (bands), ferromagnetism (the spontaneous alignment of magnetic spins) and charge-ordered stripes. This is truly the stuff of condensed-matter physics.

Phase separation is a hot topic in complex magnetic-oxide science¹. But length scales matter. Macroscopic phases of ferromagnetism and antiferromagnetism (where spins spontaneously align antiparallel) might each extend over many tens of nanometres² (Fig. 1c), whereas the charge-ordered stripes that occur in certain manganites represent

*International Workshop on Spin, Charge and Orbital Ordering in Complex Magnetic Oxides, Dubna, Russia, 1–3 July 1999.

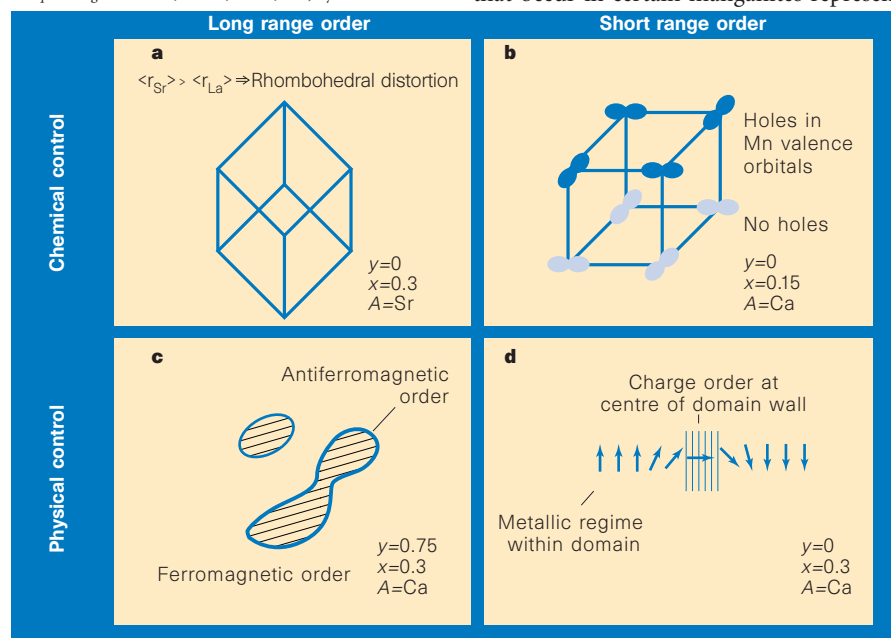


Figure 1 Long-range and short-range order may be significantly varied in the (La_{1-x}Pr_x)_{1-y}A_yMnO₃ system by chemical and physical means. a, Chemical strain is introduced by replacing La with larger Sr atoms, producing a rhombohedral lattice distortion. b, Varying the ratio of La:Ca, x, leads to an unusual layered structure when x = 0.15, with every other microscopic layer containing no charge carriers. c, Physical control is demonstrated by applying a magnetic field B in the range 0.1 Tesla < B < 1 Tesla to produce a mixture of long-range ferromagnetic and antiferromagnetic order. d, A thin-film device can be used to impose an artificial magnetic twist in a continuous ferromagnetic crystal.

microscopic phases that span just one or two nanometres³.

A new microscopic phase was proposed to explain the mysterious 15% doped Ca area of the $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$ phase diagram (that is, when $x = 0.15$). Charge-ordered antiferromagnetic ground states are the norm⁴ when trivalent La^{3+} is partially substituted with divalent Ca^{2+} in the range $0 < x < 1$ to create potential charge carriers in the Mn–O bonds. But in a wide window centred near the 33% doped Ca area ($x = 0.33$), the ground state is metallic, ferromagnetic and integral to the colossal-magnetoresistance effect, where an applied magnetic field produces a large change in electrical resistance⁵.

Interpolating the phase diagram suggests a homogeneous canted antiferromagnetic ground state around $x = 0.15$. But this turns out to be theoretically unstable if orbital dependences (D. I. Khomskii, Groningen Univ.), local spin texture (W. Koshibae, Tohoku Univ.) or compressibility considerations (M. Y. Kagan, Kapitza Institute for Physical Problems, Moscow) are combined with the 'double exchange' model⁶, in which ferromagnetic order stabilizes the metallic state.

In fact, a novel, layered structure develops below 175 K (M. V. Lobanov, Moscow State Univ.). Bond lengths of the Mn–O sublattice were accurately determined by high-resolution neutron-scattering experiments. Bond valence calculations revealed that half of the layers contain Mn^{3+} ions, whereas every other layer comprises $\text{Mn}^{3+}/\text{Mn}^{4+}$ ions and therefore contains holes — positive charge carriers (Fig. 1b). Orbital order prevents good conductivity (with a high resistivity of $0.2 \Omega \text{ cm}$ as $T \rightarrow 0$). Nonetheless, these stripes are geometrically very different to insulating ones, such as those seen at $x = 0.67$ (ref. 3). Crucially, chemical phase separation was ruled out because the neutron-scattering cross-sections for La and Ca are very different, and so the holes are not trivially confined to the Ca^{2+} -rich regions.

Unusual short-range order described in other complex magnetic oxides includes the exotic pattern seen in NaV_2O_5 below 35 K, where spins are distributed about the 'rungs of the ladder' created by V–O–V bonds (A. N. Vasil'ev, Moscow State Univ.). In this material the electrical conductivity is very sensitive to the absence of sodium (1,000% increase for every 0.1% Na removed), suggesting the possibility of interesting magnetoresistive effects, such as those now used in disk-drive heads.

What about macroscopic phase separation? The $(\text{La}_{1-y}\text{Pr}_y)_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ system was probed with neutrons, to investigate the antiferromagnetic correlations that appear on the ferromagnetic background when distortions are introduced by replacing La with the smaller but electrically similar Pr atoms (A. M. Balagurov and D. V. Sheptyakov, Joint

Institute for Nuclear Research, Dubna; Fig. 1c). Diffraction peaks representing ferromagnetic and antiferromagnetic order were found to exhibit radically different dependences on temperature, magnetic field, oxygen-isotope substitutions and the La:Pr ratio, y . This is clear evidence for two phases coexisting in the same material. Again, chemical phase separation was ruled out by a combination of techniques.

Support for macroscopic coexistence in certain manganites came from measurements revealing that the amount of ferromagnetic material present depends on the thermal history. This observation, which can be explained in terms of a charge-carrier redistribution, was made using a.c. magnetic susceptibility measurements (N. A. Babushkina, Kurchatov Institute, Moscow), and optical probes of both the magnetization and the itinerant charge carrier density (E. A. Gan'shina, Moscow State Univ.). Further evidence came from relaxation effects seen in magnetic and electrical transport data (L. M. Fisher and I. F. Voloshin, All-Russia Elec. Eng. Inst., Moscow).

Viewing the manganites through a chemical 'window' permits properties of interest to be controlled accurately (J. P. Attfield, Cambridge Univ.). Disorder strongly affects magnetic and electronic properties^{7,8}, and, in the extreme case of a physical grain boundary, large low-field magnetoresistive effects are seen^{9,10}. The common ploy of quantifying bulk disorder by average cation size $\langle r_A \rangle$ is not reliable. The correct independent variable is σ^2 , the statistical variance in the size of the different cation species present. But sometimes you can get away with $\langle r_A \rangle$. For example, although La is larger than Pr and Ca, it is not large enough to promote significant incoherent distortions in the $(\text{La}_{1-y}\text{Pr}_y)_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ system discussed earlier ($\sigma^2 \approx 0$). The science of materials stands to gain a lot from this form of analysis.

Physical control can extend beyond the macroscopic parameters of temperature, pressure and field. My own research uses thin-film devices which may, for example, be used to impose an artificial magnetic twist — that is, a magnetic domain wall (Fig. 1d). This might promote charge ordering on short length scales.

So macroscopic phase separation does indeed occur in complex magnetic oxides, but it is by no means necessarily universal. What basic picture best describes the bizarre internal configurations, distortions and phase separations across the various length scales? Are the fundamental interactions mediated by composite entities known as 'quasi particles'? One very positive suggestion was that quasi-particle excitation spectra might provide distinctive signatures and good criteria for the appropriate choice of model (A. L. Kuzemsky, Joint Institute for Nuclear Research). For example, one might expect to see spin and orbital excitations forming bound states (K. I. Kugel, Scientific Centre for Applied Problems in Electrodynamics, Moscow), and these, or other quasi-particles, might be instrumental in creating the great diversity observed. □

Neil Mathur is in the Department of Materials Science, University of Cambridge, Cambridge CB2 3QZ, UK.

e-mail: ndm12@cus.cam.ac.uk

1. Littlewood, P. *Nature* **399**, 529–531 (1999).
2. Uehara, M., Mori, S., Chen, C. H. & Cheong, S.-W. *Nature* **399**, 560–563 (1999).
3. Mori, S., Chen, C. H. & Cheong, S.-W. *Nature* **392**, 473–476 (1998).
4. Cheong, S.-W. & Hwang, H. Y. in *Colossal Magnetoresistance Oxides* (ed. Tokura, Y.) Ch. 7 (Gordon & Breach, London, 1999).
5. Ramirez, A. P. *J. Phys. Condens. Matt.* **9**, 8171–8199 (1997).
6. Zener, C. *Phys. Rev.* **82**, 403–405 (1951).
7. Rodriguez-Martinez, L. M. & Attfield, J. P. *Phys. Rev. B* **54**, R15622–R15625 (1996).
8. Attfield, J. P., Icharlanov, A. L., McAlister, J. A. & Rodriguez-Martinez, L. M. *MRS Symp.* **547**, 15–26 (1999).
9. Hwang, H. Y., Cheong, S.-W., Ong, N. P. & Batlogg, B. *Phys. Rev. Lett.* **77**, 2041–2044 (1996).
10. Gupta, A. *et al. Phys. Rev. B* **54**, R15629–R15632 (1996).

Reproductive biology

Distinguished sperm in competition

Tim Birkhead

For most species of animal, females are usually inseminated by more than one male in any particular breeding cycle. The result is competition between the sperm to fertilize the female's eggs. Such sperm competition is a powerful but subtle evolutionary force that shapes both behaviour and physiology^{1,2}. A common pattern, especially among insects and other invertebrates, is for the last of the several males that inseminate a female to fertilize most of her eggs. 'Last-male sperm precedence', as it is called, has been known in the fruitfly *Drosophila melanogaster* since 1920. But despite many investigations³, researchers have not yet

reached a consensus regarding the underlying mechanism. In their paper on page 449 of this issue, Price *et al.*⁴ describe two processes that work together in a subtle combination to produce last-male sperm precedence.

The fruitfly was one of the first animals in which sperm competition was investigated. Its various mutants provide reliable genetic markers (such as eye colour) that can be used for assigning paternity to particular males in laboratory populations. Such studies showed that, typically, the second of two males to inseminate a female fathers about 80% of her offspring. Researchers speculated about potential mechanisms. One possibi-

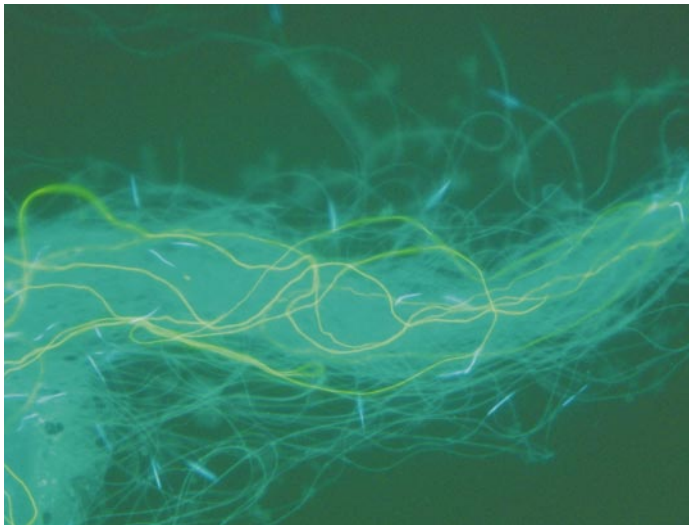


Figure 1 Untangling sperm competition in *Drosophila*. To study last-male sperm precedence, Price *et al.*⁴ used genetically modified fruitflies that produce sperm with green, fluorescing tails. These sperm can easily be distinguished from non-fluorescing sperm.

lity was that sperm from the most recent insemination physically displaces previous deposits, pushing them to the back of the female's sperm store, such that successive ejaculates could become stratified. This would result in a 'last in, first out' system⁵.

But there was no direct evidence to support the idea of stratified ejaculates, and subsequent studies revealed that last-male sperm precedence might be the result of rather more subtle processes. Some researchers, for example, have found that the seminal-fluid component of a male fruitfly's ejaculate disables the sperm from previous inseminations. Not only that, but it transpires that these ejaculatory secretions are also toxic to the female herself. So, these results showed that the more often a female copulated the sooner she died, creating a serious sexual conflict over the frequency of copulation. However, not all studies found that seminal fluid incapacitates sperm, so it has been difficult for researchers to agree on the relative importance of this mechanism⁶.

Price *et al.*⁴ have now used new techniques to assess the relative importance of displacement and incapacitation in last-male sperm precedence. In their first experiment, they allowed female fruitflies to be inseminated by two different males at intervals of either two or seven days. In both cases, as in previous studies, the second male fathered most of the offspring — 87% with a two-day interval, and a huge 96% with a seven-day interval. The authors found that displacement alone was responsible for second-male sperm precedence with a two-day interval between inseminations. But with seven days between inseminations, the second mechanism, incapacitation, worked in conjunction with displacement, resulting in almost complete precedence of the second male's sperm.

How did the authors work this out? By using a pair of males, one of which was genetically modified so that the tails of its sperm fluoresced green (Fig. 1), Price *et al.* followed the fate of individual sperm from each male

in the female's reproductive tract. This was not quite as straightforward as it sounds, because the males whose sperm fluoresced green transferred many fewer sperm than other males. But, after accounting for this effect, the authors concluded that an incoming ejaculate physically displaces sperm from one of female's three sperm stores (the seminal receptacle). So, most of the stored sperm in the female are those from the second male. Based on the distinguishable sperm, Price *et al.* also state that the sperm from different males were not stratified within the sperm receptacle, although they do not provide quantitative details.

In a second set of experiments, Price *et al.* used a spermless mutant fly to show that the incapacitation effect of seminal fluid works only when the interval between the two inseminations is seven days. This result implies that sperm must remain in the female's sperm stores for at least two days before they become vulnerable to incapacitation. The authors offer a neat explanation for this effect — they state that it prevents males

from inactivating their own sperm. In the wild, female fruitflies usually re-mate to replenish their dwindling sperm supplies after two or three days⁷. So, the interval that sperm must be stored before becoming vulnerable to debilitation is probably closer to three days than to seven. This remains to be tested.

Price and colleagues' study has taken us forward both in understanding the mechanism of last-male sperm precedence, and in identifying new tools for the job. But there is still much to be discovered. We do not, for example, know how one ejaculate physically displaces another. Nor do we know what happens to incapacitated sperm — are they merely less mobile? And why does incapacitation work only on sperm that has already been stored for several days? Comparing these results with those from other studies also raises the question of why sperm-competition mechanisms vary so much between species^{8–11}. Finally, Price and colleagues' study confirms what other sperm-competition researchers have known for some time — that males and females do not readily relinquish their reproductive secrets. □

Tim Birkhead is in the Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.

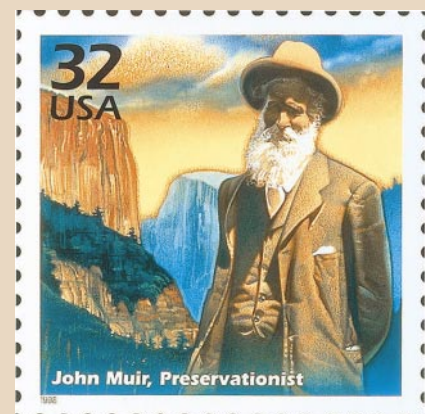
e-mail: T.R.Birkhead@sheffield.ac.uk

1. Parker, G. A. *Biol. Rev.* **45**, 525–567 (1970).
2. Birkhead, T. R. & Møller, A. P. *Sperm Competition and Sexual Selection* (Academic, London, 1998).
3. Gromko, M. H., Gilbert, D. G. & Richmond, R. C. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 371–426 (Academic, Orlando, 1984).
4. Price, S. C., Dyer, K. A. & Coyne, J. A. *Nature* **400**, 449–452 (1999).
5. Lefevre, G. & Jonsson, U. B. *Genetics* **47**, 1719–1736 (1962).
6. Gromko, M. H., Newport, M. E. A. & Kortier, M. G. *Evolution* **38**, 1273–1282 (1984).
7. Van Vianen, A. & Bijlsma, R. *Heredity* **71**, 269–276 (1993).
8. Diesel, R. J. *Zool.* **220**, 213–223 (1990).
9. Gack, C. & Peschke, K. *Zoomorphology* **114**, 227–237 (1994).
10. Simmons, L. W., Parker, G. A. & Stockley, P. *Am. Nat.* **153**, 302–314 (1999).
11. Birkhead, T. R. *Revs. Reprod.* **3**, 123–129 (1998).

Exhibition

Roots of a conservationist

John Muir, icon of Americans, and of conservationists world-wide, is being reclaimed by his native Scotland from 31 July to 2 October, in the form of an exhibition in Edinburgh to celebrate his life and achievements. In 1849, aged ten, Muir and some of his family left his native Dunbar, East Lothian, to emigrate to Canada. They became diverted to Wisconsin from where, after an initial career as an engineer and inventor, Muir was drawn to travel and map the wilds of North America. During this long period, 1870–1914, he became increasingly dismayed at the despoilation of the natural world and became the standard bearer in the battle to protect it. Over 200 sites bear his name, and he is here commemorated



on a US stamp. The exhibition is entitled "An infinite storm of beauty" and is at Edinburgh's City Art Centre. **Tim Lincoln**

Earthquakes

Slow ruptures, roaring tsunamis

Heidi Houston

Subduction zones, where one tectonic plate dives beneath another overriding plate, are the Earth's most seismically active plate boundaries, producing 90% of global seismicity as well as the largest earthquakes. Now it seems that, in these complicated zones of convergence, shallower earthquakes take much longer to 'break' than events that occur deeper at the plate interface — that is, rupture lasts much longer. Bilek and Lay (page 443 of this issue¹) analysed the durations of hundreds of earthquakes using globally recorded seismograms. Their interpretation of these inferred slower, shallow ruptures is that the ruptures propagate through material of unexpectedly low rigidity or shear modulus. The rigidity is a measure of the resistance of an elastic solid to distortion by shear (sideways or twisting motions), and is the ratio of shear stress to shear strain.

The rigidity of the rock in which a seismic rupture occurs controls the amount of slip in an earthquake of a given size, and so should exert a strong influence on the size of tsunamis generated by an underwater rupture (tsunamis are large, destructive ocean waves triggered by movement of the sea floor caused by underwater fault ruptures, landslides or volcanic eruptions; Fig. 1). The rigidity also provides information needed to analyse seismograms properly. Earthquake size is usually measured by seismic moment, which is the product of rigidity, the surface area that ruptures, and offset (or slip) between the two sides of the fault. For a lower rigidity, an underwater rupture of a given seismic moment can generate more slip or surface area of slippage, and thus a larger tsunami. So constraints on rigidity are useful for researchers who calculate the size of tsunamis.

The degree to which rigidity increases with depth also constrains the way in which subduction-zone sediments are transformed by increasing pressure and temperature from watery muds to hard metamorphosed rock. Water-laden sediments that have been produced by erosion on an adjacent continent, or scraped off the downgoing lithospheric plate, collect above the plate interface in a wedge-shaped accretionary prism (see Fig. 4 on page 446). Such unconsolidated sediments may have very low rigidities, and are probably distributed in a highly heterogeneous fashion.

Bilek and Lay's finding¹ is consistent with an emerging consensus that tsunami earthquakes (those that produce large tsunamis for their seismic moments) tend to occur in low-rigidity sediments and have a slow rupture velocity²⁻⁴. The basic technique used by the authors is to 'invert' several seismograms



Figure 1 Earth and ocean power. Aftermath of the 1992 tsunami in Nicaragua.

from each earthquake for the event depth and for source time function, a time series that describes the temporal release of seismic moment. The durations of rupture can be inferred from the source time functions, although there is some ambiguity here because the functions obtained in this way often lack a clear termination, due to noise and to a gradual rather than abrupt decrease in moment release. Other factors being equal, bigger earthquakes will have longer durations; a scaling procedure is therefore applied to remove this fairly well-defined effect so that durations and time functions from events of different size can be compared⁵.

Bilek and Lay looked at data for six subduction zones, and found a decrease in scaled rupture duration of factors of two to three as depth increases from 5 to about 30 km (see Fig. 1, page 443). The changes in duration can be interpreted on the assumption that they are entirely due to changes in the velocity at which ruptures propagate, which is controlled by the velocity at which shear waves propagate. Shear-wave velocity, in turn, is proportional to the square root of rigidity. So the long scaled durations for the shallowest quakes could imply a rigidity about five times smaller than predicted by a standard model. Another end-member interpretation is that the rigidity could follow the model, but the ratio of fault length to offset could increase markedly at shallow depths, producing ruptures of longer duration for a given moment. In this case the stress drop (which is proportional to the ratio of fault offset to length) would be much lower for shallow events.

Actually, as Bilek and Lay point out, low rigidity and low stress drop are probably coupled and probably both contribute to the very long durations they see at shallow depths. Indeed, the lower breaking strength expected from unconsolidated sediments provides an upper bound on the stress drop. Several other possible effects could compli-

cate interpretation of the durations — for instance, the shallowest ruptures might propagate in only one direction, whereas deeper ruptures may tend to expand in more than one direction.

Some further potential ambiguities in Bilek and Lay's analysis seem unavoidable. One stems from the difficulty in determining seismic moment for near-horizontal faults. Another is that a very low rigidity would also affect the estimation of the seismic moment of the earthquake^{6,7}, leading to an inevitable inconsistency in the analysis. The earthquake moments used to scale the durations were obtained from a catalogue that assumed a standard model for rigidity. If the rigidity is actually much smaller, the moments may be quite inaccurate. Correcting for this is difficult. The estimation of seismic moment depends on the detailed distribution of low-rigidity material, which is likely to be heterogeneous in these regions of rapid plate convergence. This uncertainty may reduce the constraints on rigidity. Finally, there is the well-known difficulty of determining the depths of underwater ruptures that occur close to Earth's surface; for such events, seismic waves that reflect off the nearby Earth–water interface or that bounce in the water column obscure the waves arriving directly from the source.

Workers in this field had long noticed that 'low frequency' earthquakes with anomalously low levels of high-frequency radiation and long durations seem to occur much more commonly in an offshore belt just landward of the Japan trench⁸⁻¹⁰. Such events may have shallow depths⁹ and sometimes generate larger tsunamis than expected⁸. These low-frequency earthquakes seem qualitatively similar to the shallow, long-duration events identified by Bilek and Lay, and were also inferred to occur in unconsolidated sediments⁹. Furthermore, the anomalously low rupture velocity and stress drop of the 1992 Nicaragua tsunami earthquake⁴ are roughly quantitatively consistent with Bilek and Lay's surprisingly large factor of a two-to-three decrease in scaled duration with increasing depth.

All in all, then, it seems that some earthquakes rupture slowly because the rupture is occurring in shallow, unconsolidated sediments. But this view will have rock mechanicians scrambling to explain an apparent contradiction — how can such sediments, which are thought to show 'velocity strengthening' frictional behaviour¹¹ (in which friction increases with greater sliding velocity, tending to promote stable aseismic sliding) support the unstable failure represented by seismic rupture. □

Heidi Houston is in the Department of Earth and Space Sciences, University of California at Los Angeles, 595 Charles E. Young Drive East, Los Angeles, California 90095-1567, USA.
e-mail: heidi@moho.ess.ucla.edu

1. Bilek, S. L. & Lay, T. *Nature* **400**, 443–446 (1999).
2. Fukao, Y. *J. Geophys. Res.* **84**, 2303–2314 (1979).
3. Satake, K. *Geophys. Res. Lett.* **21**, 2519–2522 (1994).
4. Kikuchi, M. & Kanamori, H. *Pure Appl. Geophys.* **144**, 441–453 (1995).
5. Houston, H., Benz, H. & Vidale, J. E. *J. Geophys. Res.* **103**, 29895–29913 (1998).
6. Woodhouse, J. H. *Geophys. Res. Lett.* **8**, 1129–1131 (1979).
7. Heaton, T. H. & Heaton, R. E. *Bull. Seismol. Soc. Am.* **79**, 813–841 (1989).
8. Wadati, K. *Kishoshushi (J. Meteorol. Soc. Japan) Ser. 2* **6**, 1–43 (1928).
9. Fukao, Y. & Kanjo, K. *Tectonophysics* **67**, 153–162 (1980).
10. Utsu, T. *J. Phys. Earth* **28**, 361–384 (1980).
11. Scholz, C. H. *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, 1990).

Apoptosis

All's well that ends dead

Timothy S. Zheng and Richard A. Flavell

Too much of a good thing can be bad. Although expanded populations of lymphoid cells are a crucial defence against infectious agents, the immune system must then eliminate the excess cells after the pathogen has been destroyed. This depends on a cell-suicide process called apoptosis, and failure to destroy the surplus lymphoid cells leads the immune system to react against its own tissues (autoimmune effects)¹. Reporting in *Cell*, Wang and colleagues² describe a hitherto unknown type of autoimmune lymphoproliferative syndrome (ALPS) that arises from congenital mutations in caspase-10, a death protease that needs to be activated for apoptosis in mammals. The discovery not only expands our understanding of the genetic alterations that underlie ALPS, but also provides the first example of caspase mutations in human disease.

Caspases are the central executioners of mammalian apoptosis. They are produced in an enzymatically inactive form, and are activated in response to apoptotic stimuli. Active caspases degrade cellular components, resulting in eventual cell death³. Given that apoptosis is critical in maintaining proper cell physiology, disruption of the apoptotic machinery is thought to contribute to various diseases, ranging from neurodegenerative disorders to tumorigenesis⁴. So, it has long been suspected that mutations in caspases might be associated with certain inherited human conditions.

First recognized by Canale and Smith more than 30 years ago⁵, ALPS is a rare, lymphoproliferative syndrome that affects people mainly during early childhood. Patients accumulate an unusual population of T cells that bear neither CD4 nor CD8, and they often develop autoimmune disorders such as haemolytic anaemia and thrombocytopenia, owing to circulating antibodies against their own red blood cells and platelets, respectively. These abnormalities resemble those observed in two autoimmune mouse models, *lpr* and *gld*, that carry autosomal recessive mutations in the death receptor Fas or in its ligand, FasL, respectively⁶.

Such similarities prompted several groups to search for potential alterations in the *Fas* and *FasL* genes in patients with ALPS, and mutations were soon discovered. The only

known mutation in *FasL* causes type Ib ALPS, and presumably results in defective binding to the Fas receptor. More than 20 mutations have been found in the *Fas* gene (causing type Ia ALPS), resulting in defective expression of, or signalling by, Fas⁷. Engagement of the Fas receptor by FasL is the first step in one of the best-studied cell-death pathways. Binding leads to recruitment of the adaptor protein FADD/MORT1, and subsequent activation of caspase-8 (Fig. 1)⁸. So, the mutations in *Fas*, which are found in both the extra- and intracellular domains, are thought to affect the binding of Fas either to its ligand, FasL, or to its adaptor FADD/MORT1. But in a number of cases, designated type II ALPS, patients have a deficiency in Fas apoptosis without mutations in either *Fas* or *FasL*. This indicates that additional modifying factors affect Fas-mediated signalling in these patients^{9,10}.

To understand the molecular basis for

type II ALPS, Wang *et al.*² traced along the Fas signal-transduction pathway of two families with the disease. Although they found no molecular abnormalities in *Fas*, *FasL*, the tumour-necrosis factor receptor-1 (TNFR1), TNFR2, FADD and caspase-8, in each family the authors discovered a distinct, germline, amino-acid substitution in caspase-10. Wang and colleagues also found that bacterially expressed caspase-10 mutants had significantly reduced enzymatic and autoproteolytic activities, and that, when transfected into mammalian cells, these mutants had a reduced ability to induce apoptosis. These results indicate that caspase-10 is defective in these ALPS patients.

Why should mutations in caspase-10 result in such a dramatic impairment of apoptosis in lymphocytes, when caspase-8 has always been considered the upstream caspase in the Fas pathway? Using a biochemical approach, the authors showed that at least one of the mutant alleles could be directly recruited into the Fas-receptor complex, together with endogenous FADD and caspase-8. The mutant allele interferes with induction of death by several death receptors, including Fas, TNFR1, death receptor (DR)-3, DR4 and DR5. So, caspase-10 mutants probably act by disrupting proper formation of the death-inducing complex. This explanation might also account for the dominant effect of the heterozygous caspase-10 mutation seen in one of the type II ALPS families.

The dominant effect seems to be respon-

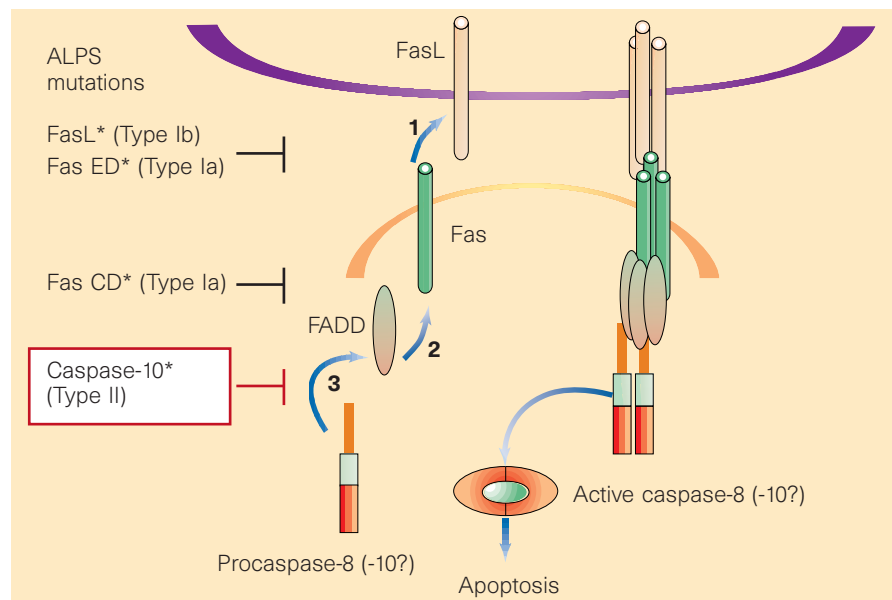


Figure 1 Defective Fas signalling pathways underlie autoimmune lymphoproliferative syndrome (ALPS). It has previously been shown that, whereas mutations in Fas ligand (FasL) and the extracellular domain of Fas (Fas ED) lead to deficient binding of the Fas receptor by FasL (step 1), mutations in cytoplasmic domain of Fas (Fas CD) probably interfere with its binding to FADD (step 2). Now, Wang *et al.*² have isolated mutations in caspase-10 from patients with type II ALPS. Recruitment of mutant caspase-10 with defective proteolytic activity into the Fas-receptor complex suggests that mutations in caspase-10 probably inhibit Fas signalling by disrupting normal recruitment and processing of caspase-8 or caspase-10 (step 3). So far, mutations in neither FADD nor caspase-8 have been found in people with ALPS, probably reflecting the importance of these proteins in embryonic development, as suggested by studies with knockout mice¹⁴

sible for the accumulation of dendritic cells in the lymph nodes of type II — but not type I — ALPS patients². Wang *et al.* showed that killing of dendritic cells by CD4-positive T cells is mediated mainly by the tumour-necrosis factor-related apoptosis-inducing ligand (TRAIL), through its receptor DR5, instead of by the Fas–FasL interaction. But dendritic cells from patients with type II ALPS are resistant to such TRAIL-mediated apoptosis. Using retroviral transduction, the authors further showed that mutant caspase-10 could interfere with TRAIL-mediated apoptosis in normal dendritic cells.

Wang *et al.* have documented, for the first time, caspase mutations in a human disease. Although their study clearly shows that people with type II ALPS carry mutant forms of caspase-10 that can interfere with apoptotic pathways of many death receptors, it raises other issues. First, it has yet to be established conclusively that mutation in caspase-10 is solely responsible for the ALPS in these patients. Are the expression levels of FADD and caspase-8 normal? And, despite the evidence that mutations in caspase-10 are responsible for the defective apoptosis in dendritic cells, the normal expression and function of TRAIL receptors in the patients should be measured.

Second, the dominant effect of the heterozygous caspase-10 mutant remains intriguing. Although the most likely possibility is that this mutant interferes with complex formation by directly binding to caspase-8 and FADD, such a picture does not explain why patients carrying the heterozygous allele have few clinical manifestations. This difference has also been noticed among type I ALPS patients with heterozygous mutations in Fas, and it probably reflects the presence of other genetic modifying elements. In the *lpr* mouse model, for example, the animals show varied symptoms with different genetic backgrounds¹¹.

Finally, the defective apoptosis of dendritic cells in patients with type II ALPS is a significant finding. The sustained lifespan of mature dendritic cells presenting self antigens could lead to enhanced autoreactivity of lymphocytes¹². So, the involvement of TRAIL — and, as revealed here, possibly caspase-10 — in mediating dendritic cell death might provide new avenues for studying the regulation of dendritic cell lifespan, and its possible role in causing autoimmunity¹³.

Five years after caspase-1 was shown to be the mammalian homologue of the nematode death protein ced-3, more than a dozen caspases have been identified³. Knockout studies have produced surprises about the involvement of caspases in mammalian development¹⁴. This latest finding by Wang and colleagues underscores the fact that caspases are indispensable in mammalian apoptosis, and further establishes these proteases as potential therapeutic targets for clinical applications. □

Timothy S. Zheng is at Biogen Inc., 14 Cambridge Center, Cambridge, Massachusetts 02142, USA.

e-mail: timothy.zheng@biogen.com

Richard A. Flavell is in the Section of Immunobiology, Yale University School of Medicine, and the Howard Hughes Medical Institute, 310 Cedar Street, New Haven, Connecticut 06510, USA.

e-mail: fran.manzo@yale.edu

1. Abbas, A. K. *Cell* **84**, 655–657 (1996).

2. Wang, J. *et al. Cell* **98**, 47–58 (1999).

3. Thornberry, N. A. & Lazebnik, Y. *Science* **281**, 1312–1316 (1998).

4. Thompson, C. B. *Science* **267**, 1456–1462 (1995).

5. Canale, V. & Smith, C. J. *Pediatr.* **70**, 891–899 (1967).

6. Nagata, S. *J. Hum. Genet.* **43**, 2–8 (1998).

7. Straus, S. E. *et al. Ann. Intern. Med.* **130**, 591–601 (1999).

8. Ashkenzi, A. & Dixit, V. M. *Science* **281**, 1305–1308 (1998).

9. Dianzani, U. *et al. Blood* **89**, 2871–2879 (1997).

10. Sneller, M. C. *et al. Blood* **89**, 1341–1348 (1997).

11. Steinberg, A. D. *Semin. Immunol.* **6**, 55–69 (1994).

12. Ludwig, B. *et al. J. Exp. Med.* **188**, 1493–1501 (1998).

13. Banchereau, J. & Steinman, R. M. *Nature* **392**, 245–252 (1998).

14. Zheng, T. S. *et al. Cell Death Differ.* (in the press).

Science policy

The business of research

Henk F. Moed and Marc Luwel

On page 433 of this issue, Plerou *et al.*¹ provide a quantitative study of the academic research system. Such studies are essential in examining claims as to the success (or otherwise) of the system in general, and forms of funding in particular. Whether the claims come from university scientists or policy makers, anecdote is no substitute for data.

The authors find that the growth dynamics of universities in the United States resemble those of business companies. In the same way that market forces operate in business, the peer review system appears to keep competition among scientists sufficiently strong. This intriguing result may mean that there is no need to make academic research at universities more business-like than it is already.

Plerou *et al.* include five different measures of research activity in their analysis. The largest database consists of 17 years of annual research and development (R&D) expenditure in science and engineering for 719 US universities. The authors find that the growth in research activity does not depend on the size of R&D expenditure (that is, the size of the university), in the same way that businesses (whether large or small) are subject to universal growth mechanisms. They find the same pattern in their analysis of the number of papers and patents published by more than 100 US universities over a similar time period. Moreover, they see the same behaviour in growth rates for research funding in English and Canadian universities, suggesting that the broad result holds for different academic systems.

Some historical context may help here. The classical von Humboldt model of universities — carrying out pure research without any consideration of practical application — was characterized by learning through science, and the unity of research and teaching²; conducting pure research was assumed to be the most appropriate training for a job in society. After the Second World War, this system underwent gradual transformation. The concept of pure research was increasingly replaced by one of fundamental research; intended to advance our knowl-

edge of nature, but often motivated by and funded for specific technological objectives.

Besides their traditional mission of providing academic training and carrying out fundamental research, universities today are expected to help solve society's problems and strengthen economic development. In consequence, a new mode of knowledge production³ has arisen, involving interdisciplinary research with the aim of applying knowledge in more rapid and flexible ways. These trends raise the question as to whether universities can continue to contribute to long-term basic research and maintain a balance between training and knowledge production and application⁴.

Plerou and colleagues are rather cautious in drawing general conclusions from their findings. They note that some may see the business sector as a model for academic research. In these terms, the academic research system may be considered effective, and one could conclude that the funding structure of research in the United States, particularly the relatively high proportion of short-term research grants, should be a model for other countries. On the other hand, the authors suggest that the similarity in the growth dynamics of research and business output may show that the 'economization' of fundamental research has been pushed too far in some Western countries. A key factor here is that the 'time horizon' (the number of years that management looks ahead) in the business sector is now typically five years or less, whereas in fundamental research it is often believed to be much longer⁵. In this context, the findings of Plerou *et al.*¹ reflect a convergence of the time horizons of business and universities. To make fundamental contributions to science, the research programmes of high-quality groups have to continue for more than five years — longer than the average research grant.

What drawbacks and limitations are there to Plerou and colleagues' analysis? Their approach aims to characterize the system as a whole. As a result, deviations from the general pattern are not discussed, although such differences may help show

how the findings should be interpreted. Take, for instance, studies in which one or both of us participated^{6,7}. In one⁶ a mechanism known as the Matthew effect was found to influence the growth rates of research departments in the natural sciences and medicine at Flemish universities. From a management point of view, these departments are academic business units. Both in terms of attracting external funding and publication output, large departments became larger and small departments remained small, contrary to the findings of Plerou and colleagues. Moreover, an over-emphasis on short-term objectives led to a decrease in the publication output per full-time equivalent spent on research.

In the other⁷, which involved an analysis of universities in the Netherlands during 1980–96, changes in the distribution of students among universities and the outcomes of research evaluation studies seem to have been responsible for a trend towards a uniformity of publication output in the natural and life sciences. Smaller universities had a higher growth rate of publication output than larger ones. This highlights the importance of basic public funding of universities, based on student enrolments, in the Dutch academic research system. In the United States, such basic funding is considerably less important than in several Western European countries, as most research activities are funded through grants⁸. So differences between academic systems in different countries need to be taken into account.

Finally, Plerou and colleagues' analysis¹ can be taken further in two respects. First,

more sophisticated indicators of research performance are available⁹, and could be used. Second, in their study all scientific disciplines are aggregated; yet there are big differences between disciplines in the amount and origin of funding, partly due to the increasing importance of targeted research programmes in priority areas. The effects of the increasing mobility of scientists, of informal networks and of the growing links between universities and other parts of the R&D system should also be considered. Nonetheless, the observed similarities between universities and business firms will stimulate further debate and research on the effectiveness of national academic research systems. □

Henk F. Moed is at the Centre for Science and Technology Studies, Leiden University, PO Box 9555, 2300 RB Leiden, The Netherlands.

e-mail: moed@cwts.leidenuniv.nl

Marc Luwel is at the Science and Innovation Administration, Ministry of the Flemish Community, Boudewijnlaan 30, B-1000 Brussels, Belgium.

e-mail: marc.luwel@wim.vlaanderen.be

1. Plerou, V., Nunes Amaral, L. A., Gopikrishnan, P., Meyer, M. & Stanley, H. E. *Nature* **400**, 433–437 (1999).
2. Schelsky, H. *Einsamkeit und Freiheit* (Rowolt, Hamburg, 1963).
3. Gibbons, M. et al. *The New Production of Knowledge* (Sage, London, 1994).
4. OECD Group on the Science System *University Research in Transition* (OECD, Paris, 1998).
5. Roussel, P. A., Saad, K. A. & Erickson, T. J. *Third Generation R&D* (Harvard Business School, Boston, 1991).
6. Moed, H. F., Luwel, M., Houben, J. A., Spruyt, E. & Van Den Berghe, H. *Scientometrics* **43**, 231–255 (1998).
7. Moed, H. F., van Leeuwen, Th. N. & Visser, M. S. *Res. Evaluation* **8**, 60–67 (1999).
8. National Science Board *Science and Engineering Indicators* (National Science Foundation, Arlington, Virginia, 1998).
9. van Raan, A. F. J. *Handbook of Quantitative Studies of Science and Technology* (North Holland, Amsterdam, 1988).

Bioenergetics

Two phases of proton translocation

Denis L. Rousseau

The 'central dogma' of bioenergetics is the chemiosmotic theory, which states that energy stored as a proton gradient across biological membranes (the proton-motive force) is converted to useful chemical energy in the form of ATP. In eukaryotes, the proton-motive force is generated across the inner membrane of the mitochondria by harnessing the energy released from serial electron-transfer events in membrane-bound proteins. This chain culminates in the four-electron reduction of oxygen to water by an enzyme called cytochrome *c* oxidase. Associated with this oxygen chemistry, four protons are pumped across the membrane in opposition to a proton gradient¹.

The molecular mechanism by which oxygen reduction is coupled to proton translocation — the redox linkage — remains a central question. Guided by spectroscopic^{2,3}, crystallographic^{4,5} and mutagenesis⁶ results, several models for redox linkage have been proposed,

but each has its flaws. One feature that has been lacking is a quantitative temporal relationship between the chemical steps and the translocation of protons⁷. On page 480 of this issue, Mårten Wikström and co-workers⁸ propose a surprising time frame for this relationship.

To study the reaction of cytochrome *c* oxidase, all four of its redox centres (Fig. 1) are reduced, and the enzyme is exposed to oxygen. Complete turnover of the enzyme requires two phases. During the oxidative phase, four electrons are transferred to the oxygen molecule, the O–O bond is cleaved, and the enzyme becomes fully oxidized. This is followed by the reductive phase, in which the enzyme is re-reduced. Looking at the reaction as a whole, four protons (the chemical protons) are taken up from the matrix side of the inner mitochondrial membrane and four electrons from cytochrome *c* are used in the formation of water at the catalytic site. Four more protons (the pumped pro-

tons) are translocated from the matrix to the cytosolic side of the membrane. In sum, then, eight charges are translocated across the dielectric barrier in the protein.

It was thought that, to efficiently capture the energy released by the oxygen chemistry for proton translocation, the proton-pumping steps occur during the oxidative phase. Indeed, ten years ago, Wikström postulated⁹, from equilibrium measurements, that the translocation of all four pumped protons was coupled to two specific reactions within the oxidative phase. This model, which has been the foundation for our thinking about proton translocation ever since, was called into question last year by Hartmut Michel¹⁰, who proposed that one proton is pumped during the initial reduction of the enzyme, before the reaction with oxygen, and that the other three protons are pumped during the oxidative phase. But Michel's model was not substantiated by experiment.

By making two types of measurement, Wikström and co-workers⁸ have now determined the amount of charge that is translocated during each phase of the enzymatic cycle. In the first set of experiments, the authors measured the pH change of liposomes with cytochrome *c* oxidase embedded in their membranes during the two halves of the catalytic cycle. In the second set, the time-resolved potential across a membrane containing cytochrome *c* oxidase was followed during turnover. It came as a complete surprise to find that the translocation of charge is incomplete at the end of the oxidative phase.

So what happens to the remaining charge? This, it turns out, is translocated only when the reductive phase immediately follows the oxidative phase. About half of the charge is translocated during each of the phases, indicating that part of the energy generated in the oxidative phase is conserved in the protein and indirectly coupled to proton translocation during the reductive phase. This remarkable result requires a shift in how we must think about proton pumping in cytochrome *c* oxidase, because it invalidates the hypothesis that proton translocation is directly coupled to the oxygen chemistry during the oxidative phase. But where is the energy stored at the end of the oxidative phase? And how is it converted to proton translocation in the subsequent reductive phase?

At the end of the oxidative phase, the four metal centres of the enzyme are fully oxidized. This state, termed O~, is a metastable intermediate that lasts for several seconds under the conditions used by Wikström and colleagues. The structure of O~ must be very different from that of the resting oxidized enzyme — the 'O' state — because reduction of O does not induce proton translocation⁸. The energy for driving the delayed charge translocation could be stored locally in a few highly strained chemical bonds near or at the catalytic site. On reduction of the metal

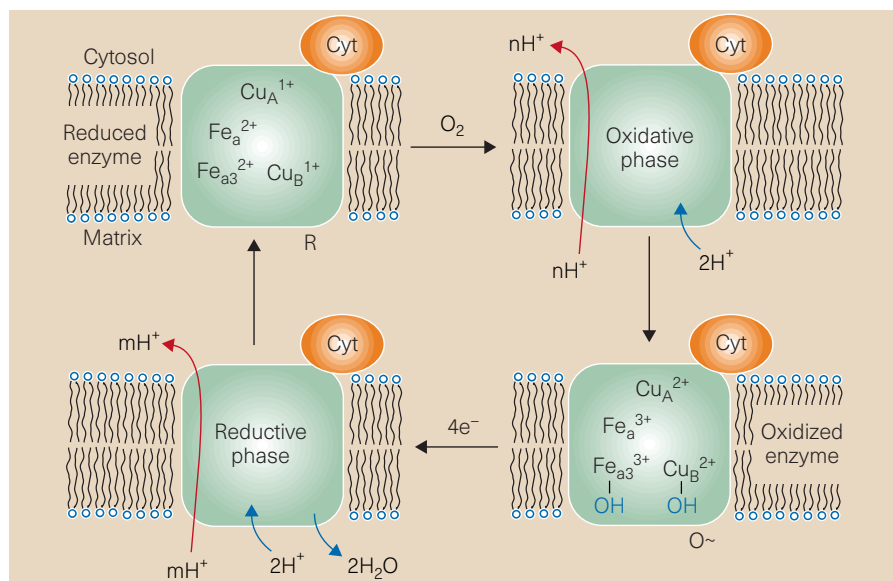


Figure 1 Proton pumping in cytochrome *c* oxidase, based on the results of Wikström and colleagues⁸. The catalytic reaction, which occurs at the Fe_{a3} – Cu_B binuclear centre, is initiated by presenting di-oxygen to the fully reduced enzyme (R). During the oxidative phase, the O_2 molecule is cleaved by accepting two protons from the matrix side of the membrane and four electrons from the metal centres of the enzyme. At the end of this phase the enzyme is fully oxidized. About half of the proton translocation occurs during this phase. Some of the chemical energy released by the oxygen chemistry is conserved in a metastable conformation of the oxidized enzyme (the $\text{O}^\sim$ state). If the subsequent reductive phase immediately follows the oxidative phase, the energy stored in the oxidized enzyme is released by completing the proton translocation. At the same time, two more protons are taken up from the matrix side of the membrane for formation of the two water molecules.

centres, this energy could be released through the charge translocation. The obvious candidates for such bonds are those formed by the coordination of exogenous ligands in the catalytic site (hydroxides, for example) with the iron and the copper in the $\text{O}^\sim$ state^{2,11}. Or, they could be those formed by the intrinsic amino-acid residues that link the redox centres to the polypeptide. Indeed, on reduction of a closely related bacterial oxidase, one of the histidine bonds that coordinates to the copper atom in the catalytic site has already been seen to rupture¹². This suggests that bond breakage or formation could be involved in the energy release.

But it is also possible that the conserved energy in the $\text{O}^\sim$ intermediate is distributed over many polypeptide bonds in the protein. Relaxation of the intermediate's conformation can be coupled to proton translocation on reduction of the metal centres. If this is the case, we need to identify the critical element that triggers release of the energy, allowing the charge translocation to occur.

One interesting possibility comes from consideration of haemoglobin. Here, binding of a ligand to the distal side of the haem moiety is coupled to protein conformational changes through the proximal histidine–iron bond that links the haem to the polypeptide. This cooperative interaction is detected in the resonance Raman spectrum of the transient species generated by photodissociation of carbon-monoxide-bound haemoglobin¹³. Changes similar to those

seen in transient spectra of haemoglobin have already been detected in the spectra of photodissociated carbon-monoxide-bound cytochrome *c* oxidase^{14,15}. It is conceivable that, on reduction of the enzyme, release of hydroxide from the haem moiety could trigger movement of the iron–histidine bond, thereby inducing the global conformational changes that are required for translocating the protons. If this process does turn out to trigger proton translocation, it is remarkable that nature could use nearly identical communication mechanisms in such vastly different proteins. □

Denis L. Rousseau is in the Department of Physiology and Biophysics, Albert Einstein College of Medicine, Bronx, New York 10461, USA.
e-mail: rousseau@aecom.yu.edu

1. Babcock, G. T. & Wikström, M. *Nature* **356**, 301–309 (1992).
2. Han, S., Ching, Y.-c. & Rousseau, D. L. *Nature* **348**, 89–90 (1990).
3. Ferguson-Miller, S. & Babcock, G. T. *Chem. Rev.* **96**, 2889–2907 (1996).
4. Tsukihara, T. *et al.* *Science* **269**, 1069–1074 (1995).
5. Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. *Nature* **376**, 660–669 (1995).
6. Gennis, R. B. *Biochim. Biophys. Acta* **1365**, 241–248 (1998).
7. Gennis, R. B. *Proc. Natl Acad. Sci. USA* **95**, 12747–12749 (1998).
8. Verhovskiy, M. I., Jasaitis, A., Verhovskaya, M. L., Morgan, J. E. & Wikström, M. *Nature* **400**, 480–483 (1999).
9. Wikström, M. *Nature* **338**, 776–778 (1989).
10. Michel, H. *Proc. Natl Acad. Sci. USA* **95**, 12819–12824 (1998).
11. Mitchell, R., Mitchell, P. & Rich, P. R. *Biochim. Biophys. Acta* **1101**, 188–191 (1992).
12. Osborne, J. P. *et al.* *Biochemistry* **38**, 4526–4532 (1999).
13. Friedman, J. M. *Science* **228**, 1273–1280 (1985).
14. Sassaroli, M., Ching, Y.-c., Argade, P. V. & Rousseau, D. L. *Biochemistry* **27**, 2496–2502 (1988).
15. Findsen, E. W., Centeno, J., Babcock, G. T. & Ondrias, M. R. *J. Am. Chem. Soc.* **109**, 5367–5372 (1987).

Daedalus

No difference at all

Sexual equality is a hot topic nowadays. Men and women are held to be identical in all respects; or if not, they should be. Bertold Brecht remarked, of the implacable social engineering of the old East German regime: “Such a government should dissolve the people, and appoint another.” This idealistic aim, says Daedalus, is now feasible, at least in the sexual sphere.

The ‘default setting’ of a fetus is female. If a male Y chromosome is present, however, it stimulates the production of fetal testosterone, which sets development along the male path. In particular, this hormone powerfully influences the growing brain. Thus, pregnant women with toxemia used to be given doses of testosterone, to counter the condition. The girls born to such women often behaved in male ways, being aggressive, unromantic, and with little interest in babies. Conversely, pregnant women treated with supplementary oestrogens often give birth to boys whose later behaviour is typically feminine. They are often shy, or lack a proper obsession with sport or technology. Much depends on the timing and the dose of the modifying hormone.

So, says Daedalus, the way seems open to true sexual equality. Provided that the sex of the growing fetus can be determined early enough, the mother could be put on a well-judged programme of hormone treatment designed to cancel the mental sexual bias of the fetus without affecting its genital development. Feminist mothers will rush to demand the treatment. The next generation will no longer show the deplorable mental differences that oppose a just and equal society. Everyone, male or female, will be moderately aggressive, moderately sexual, moderately interested in both babies and technology, and capable of competing interchangeably in all walks of life.

The moral climate of society will improve dramatically. The demand for sex from men will at last be equalled by that from women, so prostitution will vanish. The rival art-forms of pornography and romantic fiction will merge into an intriguing new unisex romantic pornography. The long, traumatic sexual odyssey of so many adolescents will also be eliminated, for both sexes will share the same outlook and aspirations. Marriage and the family will benefit enormously from this new harmony. And the old line, “My wife (or husband) doesn’t understand me”, will be totally discredited.

David Jones

Obituary

George Brown (1920–99)

Political champion of science in the United States

‘Statesman of science’ and ‘Mr Science’ were the honorifics often applied to George Brown, the Democrat congressman from California who died on 16 July, at the age of 79, following heart surgery. But neither of the titles, nor the official positions he held in the House of Representatives, conveyed the special role that the bear-like, cigar-chewing Brown long occupied in Washington science politics. If politicians rated battle ribbons, Brown would have earned a chestful during his 18 terms in the House. But he equally merited recognition as a homespun philosopher, unrestrained when given an opportunity to tell scientists to think less about budgets and more about social responsibility.

Several years ago, after delivering one of these homilies, Brown added, “At this point, many of you are probably rolling your eyes and wondering how someone who has spent the last 30 years enmeshed in the hardheaded world of science politics can take such utopian ramblings seriously, let alone espouse them in public”. The answer is that Brown savoured both utopian ramblings and hardheaded politics. In the 1970s, as a member of what was then the Science and Technology Committee of the House of Representatives, he helped create the Congressional Office of Technology Assessment; he then served on its governing board for two decades, and vainly tried to save it from the axe when Republicans took control of congress in 1995. After a petulant President Richard Nixon abolished the White House science office in 1973, for what Nixon perceived as political disloyalty by his science advisors, Brown was in the vanguard of a successful movement to restore the office during the Ford administration.

In 1990, Brown took the chairmanship of what is now called the Science Committee of the House of Representatives, in which position he emerged as the scientific eminence of Capitol Hill. Although sometimes hinting he was acting against his better judgement, Brown saved the space station from political oblivion on several occasions. He backed the Superconducting Super Collider (SSC) to the bitter end, but could not overcome the political backlash against the project’s multi-billion-dollar overruns at a time of government retrenchment.



In a departure from standard congressional practice, Brown declined to exploit his guardianship over science for his own electoral benefit, although he lived a precarious political existence, retaining his seat sometimes by wafer-thin margins — 996 out of 110,000 votes in 1996. His congressional district, encompassing most of San Bernadino, east of Los Angeles, was never the destination for any significant sums of federal research money. Brown piously refrained from the pork-barrel tactics employed by many of his colleagues to deliver science money to the folks back home — competitive peer review, he insisted, was the proper means for distributing precious research money.

Brown wryly observed that scientists flocked to see him in all seasons but one — election time, when beneficiaries of other congressmen would demonstrate their fealty with campaign contributions and public endorsements. But not the scientists, said Brown, pointing to their aloofness from the grubby side of politics. An exception occurred in 1992 when, still fresh in the chairmanship of the Science Committee, Brown faced another uphill re-election battle. Some 30 mandarins of science, hurriedly organized as the Friends of George Brown, raised \$40,000 for his campaign, among them Nobel prizewinner Leon Lederman, who declared: “The fact that so many of our nation’s best and brightest scientists want George Brown re-elected is a testament to Chairman Brown’s stature within the scientific community”.

At the time, the SSC, the dream-machine of Lederman and his colleagues in particle physics, was moving to its ill-fated climax on Capitol Hill.

First, and most of all, Brown was an enthusiast for science, extolling the benefits it could provide for humanity — if only scientists and politicians would shape up. Although he was a lifelong Democrat, Brown never hesitated to attack President Bill Clinton for what Brown regarded as inexcusable neglect of science. As for the Republicans, he attacked them, with or without good cause. At times, he sounded like a bellicose Luddite, warning of the “literally mindless process of technological evolution and economic expansion”. But whatever he said, the science establishment, and high-tech industry, knew that George Brown was a devoted friend on Capitol Hill.

In a Congress stacked with ideologically, and sartorially, prim legislators drawn from law and business, the rumpled-looking Brown — the oldest of the 435 members of the House of Representatives when he died — was a throwback to a grittier, more individualistic time in US politics. He was born in Holtville, California, in 1920. As a Quaker, he was initially a conscientious objector during the Second World War, and served in a Civilian Conservation Corps camp in Oregon. He also worked to mobilize public opposition to the wartime internment of Japanese-Americans. He later entered the army and served as junior officer for four years. After the war, he studied at the University of California at Los Angeles, where he received a bachelor’s degree in ‘industrial physics’, and later entered politics as a city councillor in Monterey Park, California.

Although he came to be best known for his political interests in science, he was also a strong supporter of civil rights legislation, and (as he often said with pride) he voted against every defence spending bill during the Vietnam war.

Several years ago, George Brown was called to the platform at a meeting of scientists to receive yet another award for service to science. After the applause subsided, holding aloft the engraved plaque, he commented, “I can’t say this award is entirely undeserved”. That was typical Brown — never hesitant to speak candidly to his scientific friends.

Daniel S. Greenberg

Daniel S. Greenberg is at 3736 Kanawha Street NW, Washington DC 20015, USA. He is author of *The Politics of Pure Science*, a new edition of which is published next month by the University of Chicago Press.
e-mail: 71732.3314@compuserve.com

Detonator of the population explosion

Without ammonia, there would be no inorganic fertilizers, and nearly half the world would go hungry. Of all the century's technological marvels, the Haber-Bosch process has made the most difference to our survival.

Vaclav Smil

What is the most important invention of the twentieth century? Aeroplanes, nuclear energy, space flight, television and computers will be the most common answers. Yet none of these can match the synthesis of ammonia from its elements. The world might be better off without Microsoft and CNN, and neither nuclear reactors nor space shuttles are critical to human well-being. But the world's population could not have grown from 1.6 billion in 1900 to today's six billion without the Haber-Bosch process.

Every one of us has to eat ten essential amino acids to synthesize the body proteins needed for tissue growth and maintenance. Agricultural crops and animals fed on crops supply almost nine-tenths of these amino acids in food proteins (aquatic species and animals grazing on grassland provide the rest). The yield of intensive agriculture is almost always limited by the availability of the nitrogen needed to produce these proteins.

Nitrogen comes from biofixation (by *Rhizobium* bacteria symbiotic with legumes and by cyanobacteria), from atmospheric deposition, and from the recycling of crop residues and animal manures. But these sources only add up to about half the global need: the other half must come from inorganic nitrogen fertilizers, whose synthesis was made possible by Fritz Haber's invention and Carl Bosch's ingenuity.

The synthesis of ammonia belongs to that special group of discoveries — including Edison's light bulb and the Wright brothers' flight — for which we can pinpoint the date of the decisive breakthrough. The archives of Badische Anilin- und Soda-Fabrik (BASF) in Ludwigshafen, Germany, contain a letter that Haber, a professor of physical chemistry and electrochemistry at the Technische Hochschule in Karlsruhe, sent on 3 July 1909 to the company's directors.

In it, Haber describes the events of the previous day, when two BASF chemists came to his laboratory to see the synthesis demonstrated: "Yesterday we began operating the large ammonia apparatus with gas circulation in the presence of Dr [Alwin] Mittasch, and were able to run it for about five hours without interruption. During this whole time it functioned correctly and produced liquid ammonia continuously ... All parts of the apparatus were tight and functioned well, so it was easy to conclude that the experiment could be repeated."

Twenty years later, Robert Le Rossignol,

Haber's English assistant, remembered that both Bosch and Mittasch came to witness the morning experiment. Paul Krassa, another of Haber's students, recalled the tension that tends to accompany such momentous events, and the fear of *Tücke des Objekts* (the spite of things). Indeed, one of the bolts of the high-pressure apparatus sprang during the tightening, delaying the demonstration for several hours; Bosch could not stay in the afternoon and returned to Ludwigshafen.

But BASF's leaders were reluctant to proceed with the development of a synthesis running at pressures above 10 MPa (100 atm). August Bernthsen, head of BASF's laboratories, was horrified: "One hundred atmospheres! Just yesterday we had an autoclave under a mere seven atmospheres flying into the air!" But Bosch, who ran the company's nitrogen-fixation research, was confident: "I believe it can go. I know exactly the capability

of the steel industry. It should be risked." And he did it: he solved some unprecedented engineering problems, and commercial production of ammonia began on 9 September 1913, just four years and two months after Haber's laboratory demonstration.

Today's ammonia synthesis has been improved in many details and is much more energy-efficient, but Haber and Bosch would recognize all the main features of their invention. Global output of ammonia is now about 130 million tonnes a year; four-fifths of this goes into fertilizers, of which urea is the most important. Rich countries could fertilize much less by cutting their excessive food production and by eating fewer animals — but even the most assiduous recycling of organic wastes and the widest planting of legumes could not supply enough nitrogen for land-scarce, poor and populous nations.

For several decades, virtually all the fixed nitrogen added to the fields of China, Egypt and Indonesia has come from synthetic fertilizers. When you travel in Hunan or Jiangsu, through the Nile Delta or the manicured landscapes of Java, remember that the children running around or leading docile water buffalo got their body proteins, via the urea their parents spread on the fields, from the Haber-Bosch synthesis of ammonia. Without this, almost two-fifths of the world's population would not be here — and our dependence will only increase as the global count moves from six to nine or ten billion people. □

Vaclav Smil is in the Department of Geography, University of Manitoba, Winnipeg, Canada.

When you travel in Hunan or Jiangsu, through the Nile Delta or the manicured landscapes of Java, remember that the children running around or leading docile water buffalo got their body proteins, via the urea their parents spread on the fields, from the Haber-Bosch synthesis of ammonia.



Haber (right) invented the process while Bosch brought the necessary engineering skills.

Chernobyl radioactivity persists in fish

After the accident at the Chernobyl nuclear reactor in 1986, the concentration of radioactive caesium (^{134}Cs and ^{137}Cs) in fish was expected to decline rapidly. The estimated ecological half-life (the time needed to reduce the average caesium concentration by 50%) was 0.3 to 4.6 years^{1,2}. Since 1986, we have measured radiocaesium in brown trout (*Salmo trutta*) and Arctic charr (*Salvelinus alpinus*), both of which are widely eaten in Scandinavia, in a lake contaminated by Chernobyl fallout^{3,4}. We have measured radiocaesium in nearly 4,000 fish, taking samples 2–4 times every year from spring to autumn. We find that the decline in radiocaesium was initially rapid for 3–4 years and was then much slower. About 10% of the initial peak radioactivity declines with an ecological half-life of as long as 8–22 years.

The concentration of ^{137}Cs , the long-lived radioactive caesium isotope with a physical half-life of 30.2 years, peaked in 1986. The radioactivity was three times higher in brown trout than in Arctic charr (geometric means: 10,468 and 3,097 Bq kg⁻¹). The decline in ^{137}Cs from its maximum in 1986 to 1998 is modelled by single- and two-component decay functions: $Q_t = Q_0 e^{-kt}$ and $Q_t = Q_1 e^{-k_1 t} + Q_2 e^{-k_2 t}$, where Q is the caesium concentration, k is the decay rate and t is time in years after the peak. The ecological half-lives are $\ln 2/k$, and are an indication of how long it will take the fish to rid themselves of radioactivity. The proportional contribution of the maximum radioactivity with slow decay rates was estimated as $Q_2/(Q_1 + Q_2)$.

The decline in ^{137}Cs was rapid during the first three (brown trout) and four (Arctic charr) years, and was then slower. Based on the initial rapid decline, ecological half-lives were estimated using a single-component decay function at 1.0 and 1.5 years for brown trout and Arctic charr, respectively, as in other post-Chernobyl studies^{1,2}, but this underestimates the time that ^{137}Cs persists in the fish. A two-component decay function

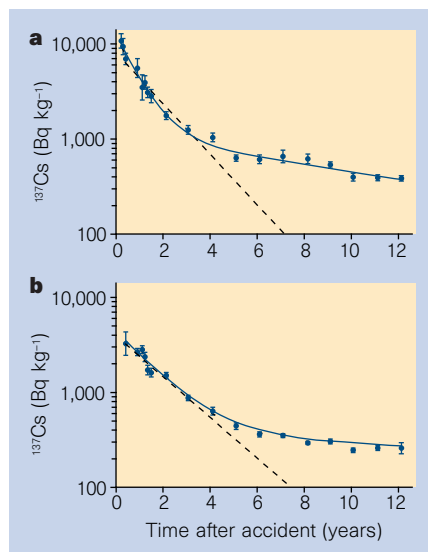


Figure 1 ^{137}Cs concentrations in fish in the study lake from peak levels in 1986 through to 1998. **a**, Brown trout; **b**, Arctic charr. Observed (points, $\pm 95\%$ confidence limits) and predicted (solid lines) ^{137}Cs concentrations are shown. Early predictions of ^{137}Cs decline based on data from peak levels until 1989 are shown for comparison (broken lines).

gives better model fits (extra sum of squares test, $P < 0.001$) than single-component models. The two-component models explain 90% and 92% of the individual variance in caesium concentration in brown trout and Arctic charr, respectively. Seasonal dynamics from 1988 (ref. 5) and size dependency⁴ in caesium levels meant that modelling should be done on all young fish (aged 2 and 3 years) until 1988, and then only on spring samples. Ecological half-lives were estimated at 0.6 and 7.7 years for the first and second components for brown trout, and 1.1 and 22.4 years for Arctic charr (Fig. 1). The second component constituted 11.5% and 10.7% of the initial peak ^{137}Cs activity for brown trout and Arctic charr, respectively.

The two-component nature of the decay indicates that the fish may be affected by

two contaminant pools. The first is a rapidly declining pool caused by caesium deposited on the lake surface and washed out from the catchment before being sorbed to catchment soils⁶. The caesium in the lake declined quickly as a result of hydraulic dilution, accumulation in bottom sediments, reduced run-off from the catchments, and loss of ^{137}Cs through outflow^{6,7}. The second is a slowly declining pool of ^{137}Cs leaking from the lake catchment^{6,8} and caesium recycling within the lake⁹.

The relative importance of the secondary pools depends on catchment and lake characteristics. The ^{137}Cs concentration in the environment may approach a steady state¹⁰, declining only as a result of decay (half-life, 30 years). This may apply generally for radioactive elements entering the biogeochemical cycles, and is supported by our estimates of ecological decay for ^{137}Cs in Arctic charr and brown trout is probably due to their different ecological niches: they segregate in habitat and diet, both of which influence caesium turnover³.

Bror Jonsson*, Torbjørn Forseth†, Ola Ugedal‡

*Norwegian Institute for Nature Research, Dronningens gt 13, 0105 Oslo, Norway

e-mail: bror.jonsson@ninaosl.ninamiku.no

†Norwegian Institute for Nature Research, Tungasletta 2, 7485 Trondheim, Norway

1. Elliott, J. M., Elliott, J. A. & Hilton, J. *Ann. Limnol.* **29**, 79–98 (1993).
2. Brittain, J. E. et al. in *Environmental Impact of Radioactive Releases* 291–298 (International Atomic Energy Agency, Vienna, 1995).
3. Forseth, T., Ugedal, O., Jonsson, B., Langeland, A. & Njåstad, O. *J. Appl. Ecol.* **28**, 1053–1067 (1991).
4. Ugedal, O., Forseth, T., Jonsson, B. & Njåstad, O. *J. Appl. Ecol.* **32**, 352–361 (1995).
5. Ugedal, O., Forseth, T. & Jonsson, B. *Ecol. Appl.* **7**, 1002–1016 (1997).
6. Smith, J. T., Leonard, D. R. P., Hilton, J. & Appleby, P. G. *Health Phys.* **72**, 880–892 (1997).
7. Santschi, P. H., Bollhalder, S., Zingg, S., Luck, A. & Farrenkoten, K. *Environ. Sci. Technol.* **33**, 519–527 (1999).
8. Hilton, J., Livens, F. R., Spezzano, P. & Leonard, D. R. P. *Sci. Total Environ.* **129**, 253–266 (1993).
9. Smith, J. T. & Comans, R. N. J. *Geochim. Cosmochim. Acta* **60**, 995–1004 (1996).
10. Smith, J. T. et al. *Environ. Sci. Technol.* **33**, 49–54 (1999).

Diamond formation from mantle carbonate fluids

Analysis of inclusions¹ has shown that natural diamond forms at pressures of 5–6 GPa and temperatures in the range 900–1,400 °C. In non-metallic systems^{2–4}, diamond has been synthesized only at pressures greater than 7 GPa and temperatures of more than 1,600 °C. We find that diamond can crystallize in alkaline carbonate-fluid melts at pressures and temperatures that correspond

to those of natural diamond formation.

For diamond crystallization, we applied the multi-anvil apparatus of a 'split sphere' type⁵. Temperature was measured by thermocouples calibrated according to data for the melting of gold, silver and aluminium, and results of diamond synthesis in the nickel-carbon system. Pressure calibration was done by the standard procedures⁵. As the components of experimental charges we used Na_2CO_3 , K_2CO_3 , graphite and oxalic acid dehydrate, generating C–O–H fluid at the required pressure and temperature, and we added minute diamond crystals as seeds

to the charges. Crystallization runs were carried out in sealed platinum ampules. The results are summarized in Table 1.

For the Na_2CO_3 –C system, at a temperature of 1,420 °C, 20 hours was too short for diamond synthesis, 30 hours allowed single diamond crystals to form, and in 40 hours the number of cubo-octahedral crystals reached 500 to 600, their size increasing to 40 μm . The presence of C–O–H fluid in Na_2CO_3 greatly enhanced the formation of diamond. At 1,360 °C for 40 hours, several hundred crystallization centres appeared in the Na_2CO_3 + C–O–H fluid + C system

(Fig. 1a) but Na_2CO_3 -C diamond nucleation was not established under similar conditions. At a temperature as low as 1,150 °C for 120 hours, there was spontaneous nucleation of diamond octahedra of up to 4 μm and growth on seeds.

In the K_2CO_3 -C system, diamonds grew on the seed crystals between 1,300 and 1,420 °C, but there was no nucleation. For the $\text{K}_2\text{CO}_3 + \text{C-O-H fluid} + \text{C}$ system, after 20 and 40 hours at 1,250–1,420 °C, we also observed growth on seeds but without nucleation. However, runs of 120 hours at 1,150 °C, in the case of Na_2CO_3 with fluid as well, led to spontaneous diamond nucleation and growth on seeds (Fig. 1b). In all experiments in which fluid was added, we found metastable graphite in the form of small crystals. X-ray diffraction analysis did not reveal any decomposition of carbonates.

Diamond nucleation in the alkaline carbonate-graphite and alkaline carbonate-fluid-graphite systems, at the pressure and temperature studied, is largely determined by kinetics and takes place only in runs lasting tens of hours. The period of induction preceding diamond nucleation and growth increases as the temperature is decreased. This is the main difference between diamond synthesis in carbonate and metal melts.

Taking into account the influence of kinetics on diamond-forming processes, the established nucleation and growth temperature of 1,150 °C can hardly be supposed to be minimal, as it is lower than in the metal-graphite systems⁶. Catalytic activity in the system decreases in the sequence $\text{Na}_2\text{CO}_3 + \text{C-O-H fluid} + \text{C} > \text{K}_2\text{CO}_3 + \text{C-O-H fluid} + \text{C} > \text{Na}_2\text{CO}_3\text{-C} > \text{K}_2\text{CO}_3\text{-C}$. Diamond growth rates vary in the range 0.01–4 $\mu\text{m h}^{-1}$, depending on the

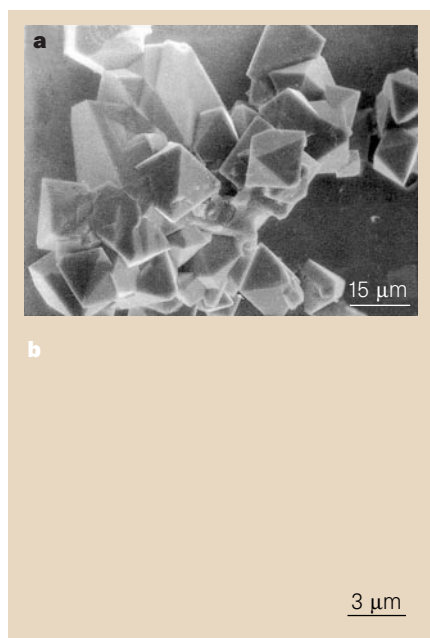


Figure 1 Scanning electron micrographs of diamonds. **a**, Octahedral diamonds (run NF-2). **b**, Diamond growth layers and spontaneous diamonds on {111} face of seed crystal (run KF-3).

temperature and composition of the crystallization medium. In the 'dry' melt of Na_2CO_3 , diamond crystallizes in the form of cubo-octahedra, and octahedra are formed in alkaline carbonate fluid-melts, which are most typical for natural diamonds.

Alkaline carbonate-fluid melts approximate the composition of a diamond-producing mantle environment^{7–9}. Considering the abundance of carbonates in diamond-bearing rocks of magmatic¹ and metamorphic¹⁰ origin, as well as the aqueous carbonaceous composition of mantle fluid⁷,

we suggest that alkaline carbonate-fluid melts represent the most likely medium for natural diamond formation.

Yu. N. Pal'yanov, A. G. Sokol, Yu. M. Borzdov, A. F. Khokhryakov, N. V. Sobolev

Institute of Mineralogy and Petrography, Siberian Branch of Russian Academy of Sciences, Novosibirsk 630090, Russian Federation
e-mail: palyanov@uiggm.nsc.ru

- Haggerty, S. E. *Nature* **320**, 34–38 (1986).
- Akaishi, S., Kanda, H. & Yamaoka, S. *J. Cryst. Growth* **104**, 578–581 (1990).
- Arima, M., Nakayama, K., Akaishi, M., Yamaoka, S. & Kanda, H. *Geology* **21**, 968–970 (1993).
- Taniguchi, T., Dobson, D., Jones, A. P., Rabe, R. & Milledge, H. J. *J. Mater. Res.* **10**, 2622–2632 (1996).
- Pal'yanov, Yu. N. *et al. Russ. Geol. Geophys.* **5**, 920–945 (1997).
- Wentorf, R. H. *Adv. High-Press. Res.* **4**, 249–281 (1974).
- Schrauder, M. & Navon, O. *Geochim. Cosmochim. Acta* **58**, 761–771 (1994).
- Harlow, G. E. *Am. Mineral.* **82**, 259–269 (1997).
- Dawson, J. B. *Nature* **195**, 1075–1076 (1962).
- Sobolev, N. V. & Shatsky, V. S. *Nature* **343**, 742–746 (1990).

Ageing, fitness and neurocognitive function

In the ageing process, neural areas^{1,2} and cognitive processes^{3,4} do not degrade uniformly. Executive control processes and the prefrontal and frontal brain regions that support them show large and disproportionate changes with age. Studies of adult animals indicate that metabolic⁵ and neurochemical⁶ functions improve with aerobic fitness. We therefore investigated whether greater aerobic fitness in adults would result in selective improvements in executive control processes, such as planning, scheduling, inhibition and working memory. Over a period of six months, we studied 124 previously sedentary adults, 60 to 75 years old, who were randomly assigned to either aerobic (walking) or anaerobic (stretching and toning) exercise. We found that those who received aerobic training showed substantial improvements in performance on tasks requiring executive control compared with anaerobically trained subjects.

Each of the 124 subjects was given a cardiorespiratory fitness test, in which the rate of oxygen consumption was measured, and a variety of cognitive tasks, including task switching⁷, response compatibility⁸ and stopping⁹. These tasks were chosen because a subset of their conditions require executive control processes and they have been shown through human lesion, neuroimaging and animal studies to be supported by frontal or prefrontal regions of the brain.

Task switching is a measure of the 'cost' of switching between tasks, indicated by the difference in reaction time between those trials in which subjects switch between tasks and those in which they continue to perform the same task; response compatibility

Table 1 Experimental results at 5.7 GPa

Number	Temperature (°C)	Time (h)	Nucleation of diamond	Growth on seeds	Thickness of diamond layer (μm)	(100) face	(111) face
Na_2CO_3 + graphite							
N-1	1,420	20	No	Yes	4	3	
N-2	1,420	30	S (5 μm)	Yes	18	20	
N-3	1,420	40	S (40 μm)	Yes	40	35	
N-3	1,360	40	No	Yes	20	6	
N-5	1,360	40	No	Yes	12	10	
N-6	1,360	40	No	No	–	–	
K_2CO_3 + graphite							
K-1	1,420	30	No	Yes	10	1	
K-2	1,420	40	No	Yes	15	3	
K-3	1,300	40	No	Yes	8	2	
K-4	1,250	40	No	No	–	–	
$\text{Na}_2\text{CO}_3 + \text{H}_2\text{C}_2\text{O}_4 \cdot 2\text{H}_2\text{O}$ + graphite							
NF-1	1,420	20	S (13 μm)	Yes	20	14	
NF-2	1,360	40	S (65 μm)	Yes	60	45	
NF-3	1,250	40	No	Yes	10	8	
NF-4	1,150	120	S (4 μm)	Yes	3	1.5	
$\text{K}_2\text{CO}_3 + \text{H}_2\text{C}_2\text{O}_4 \cdot 2\text{H}_2\text{O}$ + graphite							
KF-1	1,420	20	No	Yes	~1	~1	
KF-2	1,250	40	No	Yes	25	4	
KF-3	1,150	120	S (2 μm)	Yes	1.5	1	

S, spontaneous nucleation; the size of newly crystallized diamonds is shown in parentheses.

is a measure of the ability to ignore task-irrelevant stimuli, indicated by the difference between reaction time on response-compatible and response-incompatible trials; and stopping is a measure of the ability to abort a preprogrammed action, indicated by the reaction time to stop an action after a 'stop signal'.

Performance in other conditions on these same tasks, such as reaction time in the non-switch trials in the task-switching test, in the compatible trials on the response-compatibility task, and simple reaction time in the stopping test, depend less on executive control processes and so would not be expected to benefit from improvements in aerobic fitness.

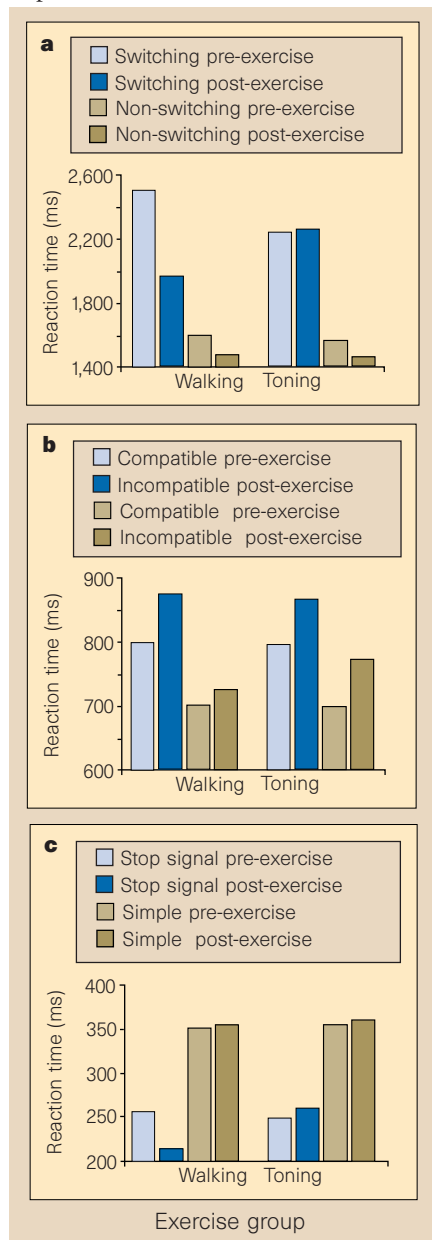


Figure 1 Tasks predicted to show selective improvements in performance for the walking but not for the toning group. **a**, Task switching; **b**, response compatibility; **c**, stopping. Experimental details are available from the authors.

Subjects in the walking group showed a significant improvement in the maximum rate of oxygen consumption (5.1%) compared with the toning group (−2.8%) after exercise training; the two groups displayed equivalent scores before exercise training.

Consistent with our 'selective improvement' hypothesis, performance improved significantly for subjects in the aerobic but not the toning group for task conditions depending on executive control processes (Fig. 1). In the task-switching test, subjects in the walking but not the toning group became much faster at switching between tasks following fitness training. Performance on the non-switch trials was equivalent for the walking and toning groups. In the response-compatibility test, the distractor-interference effect (difference between incompatible and compatible reaction times) decreased for the aerobic but not for the toning group, but there was no difference between groups on the compatible trials. In the stopping test, the reaction time for stopping was reduced for the aerobic but not for the toning subjects, whereas simple reaction time was the same for the two groups.

The three measures that were dependent on executive control processes and the integrity of the prefrontal and frontal cortex were all sensitive to the exercise intervention. However, the beneficial effect of aerobic exercise was selective: it did not affect performance on other measures in the same tasks that were not tied to frontally mediated executive functions.

The selective nature of the improvements produced by aerobic exercise, which affect only executive control processes supported by frontal and prefrontal regions of the brain, might explain the ambiguity of previous studies¹⁰ relating aerobic fitness with improved neurocognitive function. The improvement we find requires only small increases in aerobic fitness.

Arthur F. Kramer*, Sowon Hahn*, Neal J. Cohen*, Marie T. Banich*, Edward McAuley*, Catherine R. Harrison*, Julie Chason*, Eli Vakil†, Lynn Bardell*, Richard A. Boileau*, Angela Colcombe*

*Beckman Institute,

University of Illinois at Urbana-Champaign,

405 North Mathews Avenue,

Urbana, Illinois 61801, USA

e-mail: akramer@s.psych.uiuc.edu

†Department of Psychology,

Bar-Ilan University, Ramat-Gan, Israel

- Coffey, C. E. et al. *Neurology* **42**, 527–536 (1992).
- Azari, N. P. et al. *Brain Res.* **552**, 279–290 (1992).
- West, R. L. *Psychol. Bull.* **120**, 272–290 (1996).
- Kramer, A. F., Larish, J. F., Weber, T. A. & Bardell, L. in *Attention and Performance Vol. XVII* (ed. Gopher, D. & Koriat, A.) (Academic, New York, 1999).
- Black, J. E., Isaacs, K. R., Anderson, B. J., Alcantara, A. A. & Greenough, W. T. *Proc. Natl Acad. Sci. USA* **87**, 5568–5572 (1990).
- Neeper, S. A., Gomez-Pinilla, F., Choi, J. & Cotman, C. *Nature* **373**, 109 (1995).
- Rogers, R. D. et al. *Brain* **121**, 815–842 (1998).

- Rafal, R., Gershberg, F., Egly, R. & Ivry, R. *Neuropsychologia* **34**, 1197–1202 (1996).
- Hanes, D. P., Patterson, W. F. & Schall, J. D. *J. Neurophysiol.* **79**, 817–834 (1998).
- Dustman, R. E., Emmerson, R. & Shearer, D. J. *Ageing Phys. Act.* **2**, 143–181 (1994).

Developmental model for thalidomide action

Tabin has proposed a progress-zone model¹, based on published data, to explain the inhibition of limb morphogenesis by thalidomide. We do not think that his model convincingly explains the main features of thalidomide action.

We have previously listed² the factors that must be taken into account to explain the teratogenic action of thalidomide. First, the mechanism must be shown to occur in the embryo (preferably in a primate). Second, it must take into account the species specificity of thalidomide: it is teratogenic in all primates tested, has little effect in rabbits, and is not teratogenic in mice and rats. Third, thalidomide induces malformation of several organs but not of others, such as the brain. The heart, too, is an important target, and the many early postnatal deaths were probably due to cardiac malformation.

Any hypothesis on the teratogenicity of thalidomide should also be able to explain the defects induced in other organs by a common mechanism, which we believe cannot be inhibition of proliferation. It must also explain the sensitive period of thalidomide's action, which is confined to about two weeks in primates, although many induction and proliferation processes occur before and after this period. It needs to account for abnormal development but should not be too general because thalidomide has high specificity. The pattern of typical abnormalities cannot be mimicked by any other known teratogen, so its action must be unusual. Furthermore, thalidomide has many effects in different systems and dose levels, and few of these will be relevant to its teratogenic action.

With regard to Tabin's hypothesis¹, we must consider the fact that the critical period for thalidomide-induced typical limb malformations (amelia and phocomelia) is very early. According to investigations in primates^{3,4}, developmental stages 11 to 14 are affected, reaching (for example in the marmoset *Callithrix jacchus*) a maximum^{5,6} at stages 11 to 12. Upper limb buds start to develop as early as stage 11. The apical ectodermal ridge (AER), which is critical according to Tabin's hypothesis, does not occur until stage 14, too late to bring about amelia or phocomelia.

We prefer an alternative explanation of

thalidomide's teratogenic action, also based on a lack of cells in the proximal region of limb buds. An important prerequisite of limb formation in the very early stages is the migration of cells from the somatopleura (for the skeleton and connective tissue) and from the somites (for the musculature) to the sites of the prospective limb buds⁷. Inhibition of these processes by thalidomide would lead to a reduction of cells that may later react to the signals of the AER. The somatopleuric mesenchyme of the prelimb or early limb region of rhesus monkeys (*Macaca mulatta*) has already been proposed to be the site of thalidomide's action⁸.

We have suggested that thalidomide inhibits cell migration before the AER starts to function. This idea was based on the finding that thalidomide downregulates adhesion receptors, including some integrins, on the cell surface of the early primate embryo *in vivo*^{9,10}. Such adhesion receptors are not only involved in cell-cell interactions, which are known to be essential for morphogenetic differentiations in the embryo, but also take part in interactions between cells and the extracellular matrix¹¹. This could form the basis for an inhibition of cell migration. The downregulation of adhesion receptors occurs only in the primate's primordia that are susceptible to the action of thalidomide, such as the limbs and heart, but not, for example, in the brain. Furthermore, this effect on adhesion receptors occurs in primates but could not be demonstrated in rodent embryos. Our data therefore fulfil all the criteria we have listed above.

Reinhard Neubert*, **Hans-Joachim Merker†**, **Diether Neubert***

*Institute of Clinical Pharmacology and Toxicology, and †Institute of Anatomy,

Benjamin Franklin Medical Centre,
Freie Universität, 14195 Berlin, Germany
e-mail: steyn@cipmail.ukbfu-berlin.de

1. Tabin, C. J. *Nature* **396**, 322–323 (1998).
2. Neubert, R. & Neubert, D. in *Drug Toxicity in Embryonic Development* (eds Kavlock, R. J. & Daston, G. P.) 44–119 (Springer, Berlin, 1997).
3. Streeter, G. L. *Contrib. Embryol. Carneg. Inst.* **31**, 27–63 (1945).
4. Streeter, G. L. *Contrib. Embryol. Carneg. Inst.* **32**, 133–203 (1948).
5. Merker, H.-J., Heger, W., Sames, K., Stürje, H. & Neubert, D. *Arch. Toxicol.* **61**, 165–179 (1988).
6. Neubert, D., Heger, W., Merker, H.-J., Sames, K. & Meister, R. *Arch. Toxicol.* **61**, 180–191 (1988).
7. Christ, B. *et al. Anat. Embryol.* **150**, 171–186 (1977).
8. Theisen, C. T. & McGregor, E. E. *Teratology* **23**, 65A (1981).
9. Neubert, R., Hinz, N., Thiel, R. & Neubert, D. *Life Sci.* **58**, 295–316 (1996).
10. Neubert, R., Hinz, N. & Thiel, R. in *Methods in Developmental Toxicology/Biology* (eds Klug, S. & Thiel, R.) 209–226 (Blackwell, Berlin, 1997).
11. Springer, T. A. *Nature* **346**, 425–434 (1990).

Tabin replies — Neubert *et al.* suggest some useful criteria for evaluating hypotheses for the teratogenic activity of thalidomide. However, these criteria are misapplied in their

critique of my explanation of thalidomide-induced limb defects, in part because of misunderstandings regarding the current literature on limb development.

I suggested that phocomelia (missing proximal limb elements) results from lack of progress-zone proliferation in the context of continued distalization by fibroblast growth factor (FGF) signalling¹. Neubert *et al.* consider this suggestion implausible because the established period of thalidomide susceptibility in marmosets (starting at stage 11, equivalent to stage 16/17 in the chick) is before the formation of the apical ectodermal ridge (AER), the structure at the distal tip of the limb bud that is the source of FGFs. But what matters is not when a morphological AER forms, but when FGFs are first produced at the distal tip.

In all species examined, including mice, chicks and frogs, FGF-8 is first expressed in a stripe of ectoderm corresponding to the future AER just before the initial outgrowth of the nascent limb bud^{2,3}, at stage 16 in the chicken or what would be early stage 11 in the marmoset. The AER forms later, and at different relative times in different species, arising later in the mouse than in the chick, and not at all in amphibians. The stages of limb development at which marmosets are susceptible to thalidomide, as previously defined by Neubert and co-workers⁴, correspond extremely well to the stages of limb development when proximal limb elements are specified under the influence of the distal ectoderm⁵, lending explicit support to my proposal.

Neubert *et al.* also criticize my model for failing to explain other aspects of their criteria. But I never intended to address the pharmacological basis of thalidomide's effects, and indeed, as I pointed out, the pharmacology of this compound remains controversial. Whatever the mechanism, it undoubtedly has similar cellular effects in different regions of the embryo. However, in the context of the limb bud, the consequence of this is an ultimate decrease in growth of the proximal limb elements, and the point of my hypothesis was to explain in developmental terms specifically why proximal elements should be missing. This is a distinct question from the other important aspects of Neubert *et al.*'s list of criteria, such as the specificity of species susceptible to the drug or the specific set of organs affected, which probably involves factors such as differential metabolism of thalidomide.

Neubert *et al.* propose an alternative hypothesis to mine. They suggest that thalidomide could cause its characteristic limb defects by blocking cell migration. However, consideration of limb development suggests that this is unlikely because limb-bud initiation does not involve significant migration of lateral plate mesoderm

and, even if it did, blocking such migration would not lead to phocomelia.

They cite results⁶ suggesting that limb-bud formation depends on cell migration from the somatopleura and from the somites. But the work cited actually shows that there are two distinct sets of limb precursors (myogenic cells from the somites, and connective tissue or chondrogenic cells from the somatopleura) and that the myogenic cells migrate. It does not provide any evidence that somatopleural cells migrate. Indeed, subsequent work has shown that they do not⁷. Rather, the limb bud forms by local proliferation of the somatopleural mesoderm. Only the myoblasts migrate.

However, even if thalidomide were to block myoblast migration completely, it would not affect the proximodistal pattern of the limb, as removal of all myogenic precursors results in a limb with a normal skeletal pattern but no muscle⁸. Moreover, even if, despite data to the contrary, a migratory block were to decrease the number of somatopleurally derived cells as well as myogenic cells in the early limb bud, this would still not result in phocomelia. As has been shown surgically⁹ and by various methods including drug treatment, decreasing the number of cells in the early limb bud results in a narrower limb bud and subsequent loss of elements along the anteroposterior axis, but does not affect proximodistal patterning. This contrasts with the result of experimentally preventing proliferation within the progress zone during limb development, which does result in phocomelia¹⁰.

Neubert *et al.* cite the downregulation of integrin expression¹¹ to support their model. However, integrins are also important mediators of growth control and are key regulators of angiogenesis¹², so these data are compatible with other explanations for the teratogenicity of thalidomide.

Clifford J. Tabin

Department of Genetics, Harvard Medical School,
200 Longwood Avenue, Boston,
Massachusetts 02115, USA
e-mail: tabin@rascal.med.harvard.edu

1. Tabin, C. J. *Nature* **396**, 322–323 (1998).
2. Crossley, P. H., Minowada, G., MacArthur, C. A. & Martin, G. R. *Cell* **84**, 127–136 (1996).
3. Vogel, A., Rodriguez, C. & Izpisua-Belmonte, J. C. *Development* **122**, 1737–1750 (1996).
4. Neubert, D., Heger, W., Merker, H. J., Sames, K. & Meister, R. *Arch. Toxicol.* **61**, 180–191 (1988).
5. Summerbell, D. & Lewis, J. H. *J. Embryol. Exp. Morphol.* **33**, 621–643 (1975).
6. Christ, B., Jacob, H. J. & Jacob, M. *Anat. Embryol.* **150**, 171–186 (1977).
7. Wachtler, F., Christ, B. & Jacob, H. J. *Anat. Embryol.* **164**, 369–378 (1982).
8. Kieny, M. & Chevallier, A. *J. Embryol. Exp. Morphol.* **49**, 153–165 (1979).
9. Saunders, J. W. Jr *J. Exp. Zool.* **108**, 363–403 (1948).
10. Wolpert, L., Tickle, C. & Sampford, M. *J. Embryol. Exp. Morphol.* **50**, 175–193 (1979).
11. Neubert, R., Hinz, N., Thiel, R. & Neubert, D. *Life Sci.* **58**, 295–316 (1996).
12. Friedlander, M. *et al. Science* **270**, 1500–1502 (1995).

Equality in glory and culpability

Lise Meitner and the Dawn of the Nuclear Age

by Patricia Rife

Birkhäuser: 1999. 432 pp. \$39.95, £35

David Joravsky

The young men “going off to war with unbelievable enthusiasm” were “a festive joyous sight”, Lise Meitner wrote in August 1914 to Elizabeth Schiemann, another pioneering woman in the patriarchal realm of science. German women of science were galvanized by the same nationalist ideology that mobilized the country’s men of science against “the shameful spectacle of Mongols [in the Russian army] and Negroes [in the French army] being driven against the white race”. (I quote from a manifesto of 1914, signed by 93 *Kulturträger*, including Max Planck.) Meitner volunteered for front-line service as an X-ray technician, saw scientifically organized mass slaughter at first hand and still disapproved of the “naïve and strange” opinions on politics that she heard from Albert Einstein. He was one of the few scientists whose internationalism transcended physics to reveal the madness of nationalist power politics and its result.

German defeat, the consequent rise of Nazism and the greater madness of the Second World War finally opened Meitner’s mind to the ideological vision that had prompted Einstein to become a Swiss citizen even before 1914. She tried to enlighten Otto Hahn, her long-time collaborator. He had stayed in Berlin after 1938, when “non-Aryan” ethnicity obliged her to flee just as their research was approaching the discovery of nuclear fission. Hahn not only stayed, while avidly soliciting scientific advice from Meitner in Copenhagen and Stockholm, he even joined the A-bomb project of the Nazi regime. And his reaction to the pulverizing defeat of Germany was one of self-pity rather than contrition, clinging still to the nationalism that Meitner had come to see as the root of Nazism.

From exile she wrote him postwar lessons, confessing her own “great moral wrong by not leaving in ’33, because staying had the result of supporting Hitlerism”. But Hahn was no more capable of learning that lesson than he was of acknowledging Meitner’s leading role in the discovery of fission. It was he who received the Nobel prize for that, at ceremonies Meitner attended in Stockholm in 1946, stoically continuing the show of self-effacement she had adopted from the start of her entry into science.

Meitner and her nephew Otto Frisch had published the landmark report of fission in *Nature* on 11 February 1939 (Vol. 143,



Rarely has science been so entangled with ideologies of nation, careerism and lethal technology. Meitner is seen here in later years talking to students.

239–240; 1939), and should have shared the prize with Hahn, but he systematically disparaged her part in the discovery. The insult was graphically endorsed in Germany’s main science museum, which suspended a big sign — “Otto Hahn’s Workbench” — over a display of Meitner’s equipment.

Pure science has rarely been so terribly entangled with ideologies of nation, patriarchy, careerism and lethal technology as in the life of Lise Meitner. Superbly analysed by Ruth Lewin Sime in *Lise Meitner: A Life in Physics* (University of California Press), the tangle is again confronted in this “new” biography by Patricia Rife. I question whether it is new, since her copyright reports a “related biography in German” in 1990, and footnotes show that she has used some material from Sime’s book for this English version. It is a pity she did not use more, for Sime’s biography is far richer in significant information and keen interpretation. Rife’s book is shorter, but mostly lacks the clarity or narrative drive one hopes for in brevity.

But these qualities emerge occasionally, most notably in the tale of Meitner’s hair-raising flight in July 1938, when the Nazi regime was edging towards confinement and murder of Jews, and in describing the famous moment of “beautiful” revelation near Stockholm at Christmas in 1938. Meitner and Frisch, stopping on a walk in snowy woods to scribble on scrap paper, combined

Niels Bohr’s droplet vision of the atom and Einstein’s equation of energy and mass to explain why barium resulted from neutron bombardment of uranium.

Unfortunately, Rife spoils much of her book with muddled history — “Before Stalin would arrive in Washington in 1941 to press Roosevelt for Allied support ...” — and with jarring solecisms, which begin on the first page — “Like the Greek historian Herodotus pointed out” — and persist throughout. I assume she is factually wrong on Hahn’s sharing the prize money with Meitner; Sime shows that he gave some to an “assistant”, his dismissive label for Meitner.

Both biographies emphasize Meitner’s heroic isolation and inward loneliness as a woman in science when that realm was virtually forbidden to women. (However, Sime’s richer account suggests the existence of something like a sisterhood of defiant pioneers, subject, of course, to the obligatory show of womanly self-effacement and to the nationalism that overrode sisterly sentiment.)

Meitner’s lifelong friend, Elizabeth Schiemann, was dismissed from her university post in 1940, and was sufficiently anti-Nazi to shelter a Jew. But she was adamant in her postwar brush-off of Meitner’s stern new vision of German collective responsibility; as a consequence, their friendship cooled. The chemist Ida Noddack earned Meitner’s hostility with pro-Nazi sentiments in politics

even before she made her paltry claim to priority in discovering fission.

Meitner's truest *Doktorschwester*, if I may improvise on the *Doktorvater* of German academic tradition, proved to be a Swedish physicist, Eva von Bahr-Bergius, who did all she could to help Meitner in exile. But von Bahr-Bergius was as unsuccessful as Bohr, who did all he could in the backstage struggles of the 1940s to get Meitner the Nobel prize she so obviously deserved. Squalid academic politics overcame the factual record of 1938–39: Hahn was confused by his chemical findings, which were brilliantly clarified by the physics of Meitner and Frisch.

A deeper trouble than the 'woman' problem, or even the 'German' problem, attends Meitner's life in physics. Is her rightful claim to the "beautiful" discovery of fission separable from her "great moral wrong" in staying at the bench in Berlin after 1933? Those who urged her to leave earlier could not promise her the research opportunity that gave meaning to her life. These particularities express a torment common to scientists and engineers long before the Nazi regime and nuclear fission. When Hahn felt some qualms over his work on poison gas in the First World War, Meitner gave him the stock excuses: you're only following orders; others will do it if you don't; if our virtuous country does not make this or that lethal thing, the evil enemy will.

Systems that keep learned servants in line with such excuses overwhelm the exceptional individuals who rebel. James Franck was the only recruit to the Manhattan Project who asked in advance for the right to be involved in judging the use of the nuclear bomb that would come out of it. When the time came, and he produced the 'Franck report' recommending against the nuclear bombing of Japanese civilians, he and his report were brushed aside by the men in command — specialists are on tap, not on top.

In many nation states, including democracies, scientists have been crucial agents of the mechanized wholesale slaughter that has distinguished the twentieth century. Einstein noted in 1914 that "humanist" scholars took the lead in justifying that madness, but he later acknowledged the special contribution of physical scientists. "I believe," he wrote in 1946, "that the terrible decline in man's ethical behaviour is due primarily to the mechanization and depersonalization of our lives — a disastrous by-product of the development of the technological-scientific intellect. *Nostra culpa!*" (from Albrecht Fölsing's *Albert Einstein*, Penguin). Heroines like Meitner demanded equal opportunity to strive for beautiful discovery, and to share that culpability. □

David Joravsky is in the Department of History, Northwestern University, Evanston, Illinois 60208, USA.

Journey into a strange affliction

A Cursing Brain? The Histories of Tourette Syndrome

by Howard I. Kushner

Harvard University Press: 1999. 303 pp.

\$29.95, £18.50

Mary Robertson

People with Tourette syndrome can find it extremely difficult to obtain a diagnosis for their symptoms, and suffer years of misunderstanding, ostracization and pejorative nicknames. The syndrome is diagnosed by the presence for more than a year of multiple motor tics and one or more vocal tics. Coprolalia (the inappropriate uttering of obscenities) occurs in about 10 per cent of patients. Other symptoms include copying behaviour (for example, echolalia) and obsessive compulsive behaviour.

Tourette syndrome is not uncommon. Howard Kushner hints at the idea that it is a spectrum disorder and highlights its genetic background. He introduces the recent (if speculative) suggestion that it is an autoimmune disorder, and weaves the streptococcal story into the tapestry of the syndrome. It has recently been suggested that there is a subgroup of patients whose aetiology includes a group A streptococcal infection. Kushner takes us through the evolution of Tourette syndrome from a psychoanalytic to an organic disorder, and shows how its classification as a mental disease has metamorphosed, so that the diagnostic criteria and age at onset seem to change with each revision. Through the medium of Tourette syndrome, Kushner cleverly reveals the fascinating history of neurology, psychiatry, medicine and culture.

My favourite part of the book is "The Case of the Cursing Marquise", in which Kushner performs an intricate historical dissection of the Marquise de Dampierre, the most cited Tourette syndrome patient, whose story was first reported by Jean-Marc-Gaspard Itard in 1825. Itard knew the afflicted Marquise well and described her clinically. In fact, neither Georges Gilles de la Tourette nor Jean-Martin Charcot ever treated her, even though it was de la Tourette, and not Itard, who gained eponymous fame for describing such cases (including hers) in 1885.

Using an appealing intimacy, Kushner describes many of the other famous and not so famous people who may have had Tourette syndrome. These include historical figures and some of today's well-known patients, such as Sigmund Freud's patient Frau Emmy von N and Oliver Sacks' Witty Ticky Ray. For each patient, personal details and arguments for a diagnosis of Tourette syndrome are given in such a way that I felt not only that I had met the patients, but also



Gilles de la Tourette: one of the treatments for his eponymous syndrome was circumcision.

that I was able to make up my own mind on the possible diagnosis.

We are introduced to many of the clinicians, both well known and little known, who have taken a keen interest in Tourette syndrome over the years. He begins with the famous historical trio (Itard, Gilles de la Tourette and Charcot), goes on to other early French physicians, and ends with today's clinicians, including Arthur Shapiro (who died in 1995), Elaine Shapiro and Sacks — who has made Tourette syndrome a household name.

Kushner does not shy away from controversy, joining the 'battle' of the psychoanalytic versus the organic approach to Tourette syndrome that dominated the literature for years. He also discusses hotly debated topics such as the phenotype for the syndrome. Views on this range from a belief that there is no associated characteristic psychopathology, to a wide variety of psychiatric phenotypes and to the more parsimonious view taken by the group at Yale University that only obsessive compulsive behaviours/disorders and some types of attention-deficit hyperactivity disorder are genetically related to Tourette syndrome.

There has been an appalling and wide variety of treatments for the syndrome — circumcision, carbon dioxide inhalation, surgical removal of sinuses and tonsils, anti-inflammatory agents, speculative treatments today such as plasmaphoresis, and finally the (now accepted) traditional medical treatments such as haloperidol and clonidine.

My only reservation about the book is its title. *The Cursing Brain?* highlights the scatological nature of the syndrome, when there is so much more to the disorder than that. Kushner does, however, justify the title by saying that, because today Tourette's "has ... been shorn of its most well-known symptom"

(coprolalia), we are running the risk of “sanitizing” the syndrome and “ostracizing those who curse” — to whom the book pays tribute.

This book is a ‘must’ for anyone interested in the history of medicine, neurology and psychiatry as well as Tourette syndrome. There is no doubt that this is the best exposition of the syndrome’s history in the literature. The sources for the book, which make it unique, include the US Tourette Syndrome Association archives and newsletters, many taped personal interviews, letters to the author, previously untranslated articles and dissertations given to the author, as well as personal knowledge. Such details contribute to the very personal flavour of the book which is unique in a textbook of medicine. Kuschner’s is one of the most exciting and intriguing textbooks I have read: clinically and historically correct, extremely well written and an erudite and scholarly treatise. □

Mary Robertson is in the Department of Psychiatry, University College London, Wolfson Building, London W1N 8AA, UK.

Getting at where it hurts

Pain: The Science of Suffering

by Patrick D. Wall

Weidenfeld & Nicolson: 1999. 186 pp.

£12.99.

John Carmody

Scientists tend to scorn what they consider ‘popular’. Yet they generally expect the populace to support their work gratefully or, in the case of physicians, to accept their diagnoses and prescriptions credulously. When a venturesome few do write for audiences beyond their cloistered laboratories and clinics, their colleagues’ response is likely to be something akin to Samuel Johnson’s slur upon a woman’s preaching — “It is not done well; but you are surprised to find it done at all.” Still, the truth must be told, whatever the motives for such ventures, even if it may sound curmudgeonly.

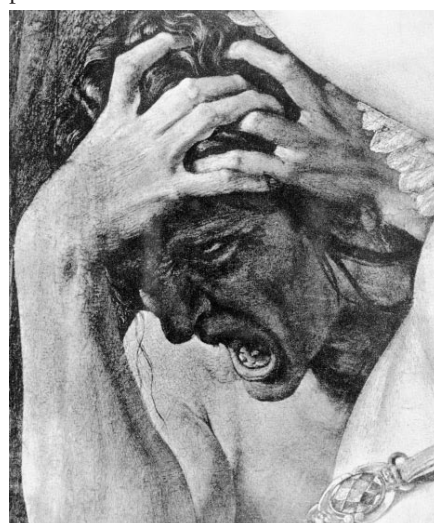
The field of pain research has flowered wonderfully in the past 30 years, in both its clinical and scientific facets, and Patrick Wall has been a highly influential figure. As co-founder and enduring co-editor of the journal *Pain*, and as co-ordinator of the so-called ‘gate control’ theory of pain (which proposed that nerve cells in the spinal cord could, depending on circumstances, operate like a gate that is either open to or closed against painful input), Wall’s achievements are genuine and substantial. As well as speaking to the profession, he has been willing to communicate with a wider public: *The Challenge of Pain* (Penguin, 1982), for example, which he wrote with his Canadian friend and colleague Ronald Melzack, remains a significant and valuable book. Whatever Wall’s

intentions, the grandiloquently titled *Pain: The Science of Suffering* is not that.

Despite its extravagant promise to “explore all we know about the nature and causes of pain”, it is a far simpler and shorter book than that earlier one, and its intended readership is not at all obvious; to a real extent, its contents are merely a *rechauffé*. It is, to be blunt, an old man’s *pensées*, and is, for all its liberal use of sub-headings, over-discursive, solipsistic and scrappily documented. I am not saying that a ‘popular’ book should be encrusted with footnotes, but it must at least provide some plausible basis for its material. Wall’s new work relies too much for my comfort on its author’s ‘introspection’, and on dogma and an appeal to the reader’s respect for the author’s gravitas. The fact that there is a scattering of careless errors (such as his statement that acupuncture “was well known in the West in Elizabethan times, beginning with a textbook ... in 1683”, or the suggestion that the Latin word *placebo* originated with Chaucer) erodes one’s confidence.

Wall’s chapter on “The philosophy of pain” is not especially lucid. Many fine minds have, over the centuries, grappled with the plethora of puzzles about pain, and I am certainly not adequate to do them justice: nor, on this evidence, is Wall — his approach being compromised by his dismissal of their efforts as a “confused mess”. He is hostile to mind-body dualists for reasons that are not cogently explained; he is hostile to Catholicism, especially its tradition of valuing suffering which, I suspect, he does not understand; he tells us that he despises something for its triviality, yet trivializes or misrepresents ideas he does not value or appear to comprehend.

He also tells us — in an otherwise interesting chapter on “The placebo response” — that, in one unspecified study, “no personality type was found diagnostic of the placebo responder”, yet only a few lines later he advises, “Avoid pessimists if you are looking for placebo reactors”. He also claims that —



Excruciation: Bronzino’s sculpture of pain.

apart from a few trivial and serendipitous (but useful) side effects — the pharmaceutical industry has, over the past 50 years, made no advances in pain treatment. Yet in that period, for example, opiates with greater potency but attenuated side effects, and far safer anaesthetic agents, have appeared.

Some of Wall’s criticisms of medical education and clinical culture are telling, even if the whole book shows an invalidating Anglo-centricity. He also proposes an interesting (though not wholly convincing) concept of pain as an action-directed experience, and writes, “Pain is then best seen as a needs state, like hunger and thirst, which are terminated by a consummatory act”. Can this be true? At the risk of my sounding Platonic, and describing pain, not as a consequence of tissue damage but as a quality of the psyche, opposite to pleasure, could much the same not be said of the concatenation of sensory inputs that determine our responses to music or painting? Those sensations, surely, warrant consideration as a special functional category, as Wall insists on exclusively for pain.

So, in the end, we will all do far better to disregard this slight book — which mostly reads as if it had been dictated to a machine — return to *The Challenge of Pain* and look to Wall’s accumulated papers (for all that some colleagues have been sceptical of their content) as a more fitting testament to his achievements. □

John Carmody is in the School of Physiology and Pharmacology, University of New South Wales, Sydney 2052, Australia.

A provincial setting for the oceans

Ecological Geography of the Sea

by Alan R. Longhurst

Academic: 1998. 398 pp. £54.95, \$79

Yves Dandonneau

Oceanographers now have a global view of the circulation of oceanic water. This view is especially well depicted and explained by Matthias Tomczak and Stuart Godfrey in their *Introduction to Regional Oceanography*, in which the mechanisms driving the currents and controlling upward and downward motions are described for non-specialists.

Alan Longhurst frequently refers to this book, and also to Harald Sverdrup’s inspired explanation of the phytoplankton spring bloom at high latitudes, which is valid over much of the ocean. Given these current global views and the author’s experience and profound knowledge of the oceans, I was prepared for a thoughtful and comprehensive description of biology in the oceans. The most interesting and original part of the book is the first six chapters, which cover all the provinces and link the large-scale physics

of the ocean to the ecosystems. Afterwards, however, Longhurst divides his book into as many chapters as there are provinces of the oceans (about 50!). The result is a long, tedious enumeration of descriptions of regions, each time with the same plan and, inevitably, often the same content. After all, every province is almost as complex as the entire ocean, and this complexity can hardly be described in three or four pages. For example, Longhurst makes the Gulf of California a subdivision of the California Current Province, but he recognizes that it could as well be divided into 14, or even 33 locations, taking it beyond the reach of this book's format.

In 1,000 years, one drop of ocean water may pass through most of the provinces. The author thinks that "the debate between partitions and a continuum is somewhat sterile because both concepts will be required for different purposes, and they are not mutually exclusive". To me, the continuum sounds much more exciting.

In reading this book, I have realized how much the generation time of grazers varies with latitude, and the consequences this has for the food chain. This alone would have made a nice chapter, but I had to pick such pieces of information from here and there throughout the 50 provinces.

The division into provinces also results in some oddities. Should we consider that the region of the Pacific Ocean that is affected by the El Niño has moving boundaries? The response time of most zooplankters in this region is much less than the duration of El Niño events, and the distribution of tuna changes during these events, so the answer should be yes. Why not consider, then, that the boundary between the North Atlantic Subtropical Gyral Province and the North Atlantic Drift Province moves similarly, with the northward shift of the summer thermocline? Indeed, many grazers in these provinces have short generation times. The boundaries proposed here are subject to endless discussion.

If this book is tedious to read, it is extremely well documented and precisely written, and this is remarkable considering the diversity of sources and sometimes the contradictions Longhurst had to work with.

In the tropical Atlantic and Pacific, my own area of expertise, I was surprised that there was no mention of the tropical instability waves in the Atlantic, and that the Guinea Dome was placed in the Western Tropical Atlantic Province while its analogue in the south, the Angola Dome, was placed in the Eastern Tropical Atlantic Province. Overall, however, the summaries that are given for these two large regions are precise and exhaustive.

Everyone will find something of interest in this book. Finally, we oceanographers like to give a name to the areas where we go on

cruises. From this point of view, the partition proposed here is certainly the best one we have today. □

Yves Dandonneau is at the Laboratoire d'Océanographie Dynamique et de Climatologie, 4 place Jussieu, 75252 Paris cedex 05, France.

Loose round-up lacks even circular logic

A History of the Circle

by Ernest Zebrowski Jr

Rutgers University Press: 1999. 214 pp. \$28

Ivor Grattan-Guinness

Popular books on or around mathematics appear quite regularly nowadays. This one nominally takes the circle as its theme, both for its mathematical properties and for its use in the physical sciences. Based upon teaching in American colleges, the level is fairly elementary.

The early chapters treat several manifestations of the circle in antiquity, including the use of rollers in transporting heavy objects and some features of astronomy. But after that the circle gives way to conic sections and closed but non-circular orbits as studied from Kepler to Newton. It comes back later rather indirectly via oscillations, pendular and simple harmonic motions and waves, and associated mathematical theories such as Fourier series. Late chapters skim across relativity theory and quantum mechanics, sandwiching an interlude on ancient architecture.

While the emphasis on applications is well taken, the overall impression is rather incoherent: links to circularity are tenuous in places, and many topics specifically tied to circles are omitted. The most obvious case concerns π . Some properties are stated, but no highlight is given to its four main roles, of relating the diameter of a circle not only to its

circumference (hence the eighteenth-century choice of the letter π , for *perimetros*) but also to its area, and to the surface area and volume of the sphere. It is enlightening, and also not at all obvious, that the same factor π is involved. The enabling theorem for the circle — area = (circumference/2) $\times$ (diameter/2) — is a profound result; the author presents one intuitive argument for it but does not register its significance. All sorts of other cultural possibilities are ignored; for example, that the classical problem in Greek geometry of squaring the circle may echo a desire, very evident with the Egyptians, of uniting the circular heavens with the Earth, long symbolized by the square.

As for history, the author states at once that he gives no conventional version. The accuracy of his statement can be fairly judged by a few typical quotations. "It was a long time before anyone had a practical need to measure angles to a precision much finer than one degree." In his *Elements*, Euclid "laboriously proved every mathematical relationship that was then known". "What Descartes did accomplish was to use rectangular co-ordinates in a new way" (he actually introduced analytic geometry). Einstein's creation of special relativity in 1905 was effected by "conduct[ing] his Gedanken experiments through mathematical logic"; indeed, in general "Scientists depend upon mathematical logic to generate their theoretical predictions" ("If only they did," Bertrand Russell might have sighed). Another circular sign comes to mind: 0 per cent for accuracy.

A good heuristic and historical introduction to parts of mathematics using the circle as its theme would make an excellent contribution to the popular understanding of mathematics. Perhaps somebody can take up the challenge. □

Ivor Grattan-Guinness is at the Middlesex University, Queensway, Enfield, Middlesex EN3 4SF, UK.

New in paperback

Countdown: A History of Space Flight

by T. A. Heppenheimer

Wiley, £13.99, \$17.95

"... The history of spaceflight is unfinished. So too is Heppenheimer's book; the publisher left the last page blank, inadvertently one assumes. Authors sue for such incompetence, but in this case it might be a fitting conclusion to this useful but necessarily inconclusive study". Alex Roland, *Nature* 392, 143–145 (1998)

The Covenant of the Wild: Why Animals Chose Domestication

by Stephen Budiansky

Yale University Press, \$14.95

"Budiansky's thesis is that human relationships with other species are not so much the 'exploitation' they are often made out to be, but

arise from quite natural coevolutionary processes like those between other pairs of species ...

Perhaps the strength of the book is that it raises issues and makes points and then doesn't tell one what conclusions to draw." Marian Stamp Dawkins, *Nature* 356, 487 (1992)

Science, Jews & Secular Culture: Studies in Mid-Twentieth-Century American Intellectual History

by David A. Hollinger

Princeton University Press, \$14.95, £12.50

Re-enchanted Science: Holism in German Culture from Wilhelm II to Hitler

by Anne Harrington

Princeton University Press, \$19.95, £11.95

Dynamic measurement of myosin light-chain-domain tilt and twist in muscle contraction

J. E. T. Corrie*, B. D. Brandmeier*, R. E. Ferguson*, D. R. Trentham*, J. Kendrick-Jones†, S. C. Hopkins‡, U. A. van der Heide‡, Y. E. Goldman‡, C. Sabido-David§, R. E. Dale§, S. Criddle§ & M. Irving§

* National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

† MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

‡ Pennsylvania Muscle Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6083, USA

§ Randall Institute, King's College London, 26–29 Drury Lane, London WC2B 5RL, UK

A new method is described for measuring motions of protein domains in their native environment on the physiological timescale. Pairs of cysteines are introduced into the domain at sites chosen from its static structure and are crosslinked by a bifunctional rhodamine. Domain orientation in a reconstituted macromolecular complex is determined by combining fluorescence polarization data from a small number of such labelled cysteine pairs. This approach bridges the gap between *in vitro* studies of protein structure and cellular studies of protein function and is used here to measure the tilt and twist of the myosin light-chain domain with respect to actin filaments in single muscle cells. The results reveal the structural basis for the lever-arm action of the light-chain domain of the myosin motor during force generation in muscle.

Although the structures of some 10,000 proteins have been described at near-atomic resolution, the conformational changes accompanying cellular function have been characterized for relatively few proteins. These motions are difficult to study using crystallography, where proteins are generally trapped in a single conformation. Protein conformations *in vivo* are dynamic, and may be influenced by binding in macromolecular complexes or to membranes. Here we describe a method for measuring protein structural changes in cells on the timescale of their physiological function. Pairs of cysteine residues are introduced into the protein at sites chosen using its three-dimensional structure and are cross-linked by a bifunctional rhodamine¹. This defines the orientation of the bifunctional rhodamine fluorescence dipole in terms of protein-structure coordinates. When the labelled protein is returned to its native environment, polarization of rhodamine fluorescence can be used to measure the orientation of the bifunctional-rhodamine dipole. The orientation of a protein domain can be calculated from observations on a suitable set of such dipoles, and the conformational changes that govern the cellular function of the protein thereby determined.

We used this approach to measure orientation changes in myosin during contraction of skeletal muscle fibres. Force generation and relative sliding between myosin and actin filaments in muscle are thought to be driven by a conformational change in the myosin head regions that bridge the two sets of filaments^{2–4}. The structures of the actin filament and myosin head have been described at atomic resolution^{5–8}, and that of the complex between them has been visualized at lower resolution by electron microscopy in the absence of nucleotide⁹ and with ADP bound at the active site of myosin¹⁰. These structural data, combined with extensive biochemical and physiological characterization of the interaction between myosin, actin and ATP, led to the 'lever-arm' model for the force-generating conformational change in the myosin head, in which its elongated light-chain domain changes orientation with respect to the actin-bound catalytic domain during the ATPase cycle^{8–15}. Recent spectroscopic^{16–18} and low-angle X-ray diffraction¹⁹ data from muscle fibres are consistent with this model, but have not characterized the three-dimensional changes in the orientation of the light-chain domain.

Here we use the bifunctional rhodamine fluorescence-polarization technique to measure these motions.

Incorporation of bifunctional rhodamine

The myosin regulatory light chain (RLC) was used to introduce bifunctional rhodamine (BR) dipoles into the light-chain domain of myosin heads in muscle, because RLC can readily be exchanged into permeabilized muscle cells in solutions of low Mg^{2+} concentration^{20,21}. The bifunctional rhodamine reagent (BR-I₂; Fig. 1a)

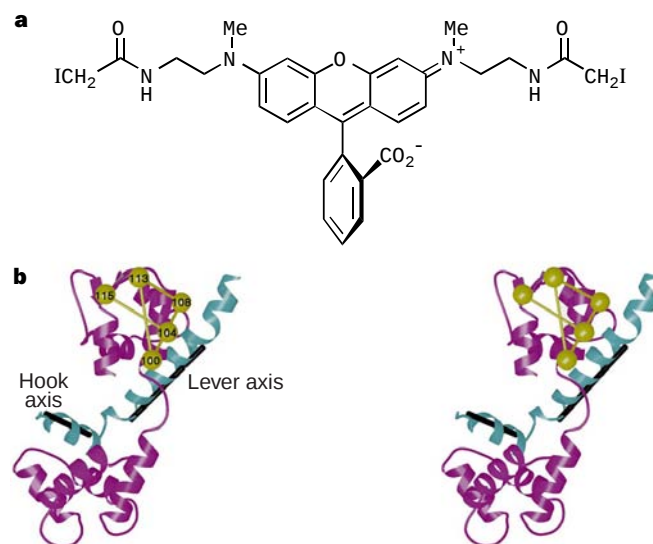


Figure 1 Bifunctional rhodamine labelling of RLC. **a**, Structure of the bifunctional rhodamine reagent (BR-I₂)¹. **b**, Stereo pair of the RLC region of the myosin head, showing the RLC (magenta), the lever and hook axes (black), part of the myosin heavy chain (blue) and the four pairs of cysteine residues that were crosslinked by bifunctional rhodamine (green spheres and rods), modelled with LOOK (Molecular Application Group, Palo Alto), using coordinates provided by I. Rayment⁷. Graphics prepared using Bobscript⁴².

contains a central rhodamine fluorophore flanked by two flexible linkers, each terminating in an iodoacetamido group¹. When these groups react with a pair of suitably placed cysteines, the average orientation of the fluorescence dipole, which is aligned with the long axis of the three coplanar rings of the fluorophore²², is expected to be parallel to a line joining the cysteines. Crystal structures of RLC bound to myosin heavy-chain fragments^{7,23} were used to select four pairs of solvent-accessible locations for the cysteines: 100–108, 100–113, 104–115 and 108–113 (Fig. 1b). These sequence numbers refer to wild-type chicken gizzard myosin RLC, which has a single cysteine at 108. New cysteines were introduced by site-directed mutagenesis, and Cys 108 was replaced by alanine in the 100–113 and 104–115 RLC mutants. Stoichiometry and specificity in labelling each pair of cysteines with BR-I₂ was confirmed by high-performance liquid chromatography, electrospray mass spectrometry and tryptic digestion. For example, the measured mass of 100–BR–108 RLC was 20,227.5 daltons (calculated 20,226.7 daltons). Limited tryptic digestion produced a fluorescent peptide, Leu 93–Arg 123 (measured mass 4,030.8 daltons; calculated, 4,031.5 daltons) containing the two linked cysteines. Prolonged digestion led to further peptide cleavage between the linked cysteines at Arg 103, with the expected mass increment of 18 daltons for hydrolysis (measured mass 4,048.6 daltons), because the two product peptides remain cross linked by bifunctional rhodamine.

Three assays showed that RLC function was unaffected by mutagenesis and bifunctional rhodamine labelling. BR–RLCs competed effectively with unlabelled wild-type RLC in competitive binding assays using myofibril suspensions. BR–RLCs replaced 40–90% of native RLC in either isolated myosin head fragments or single glycerinated muscle fibres during a 30-min incubation at 30 °C in solutions of low Mg²⁺ concentration. Most critically, Mg²⁺–ATPase rates of scallop myofibrils containing any of the four BR–RLCs were not significantly different from those with wild-type chicken gizzard RLC, in either the presence or absence of calcium. Scallop myofibrillar Mg²⁺–ATPase activity is regulated by calcium binding to the myosin essential light chain²³ but depends on the precise conformation of the RLC²⁴.

Orientation of RLC region in muscle fibres

We characterized the orientation of the RLC region of endogenous myosin heads in single skinned fibres from rabbit psoas muscle in three well-defined steady-state conditions (relaxation, active contraction and rigor), and in exogenous myosin head fragments (myosin subfragment-1, S-1) diffused into muscle fibres and bound to actin filaments in the absence of ATP. The S-1 data allow a quantitative comparison with the conformation deduced from electron microscopy⁹. The conformation of endogenous myosin heads bound to actin in the absence of ATP (in rigor) is expected to be similar to that of exogenous S-1, and may correspond to the conformation after the structural transition that generates force. By contrast, in relaxed muscle the heads are detached from actin, and their conformations should be determined by interactions within the myosin filament. In active contraction, some of the heads attach to actin and generate force^{16,25–27}. The conformations and motions of these force-generating heads are the essential structural determinants of the mechanism of muscle contraction.

Three independent fluorescence polarization ratios, $P_{||}$, $^xP_{\perp}$ and $^yP_{\perp}$ (see Methods and refs 16, 28) were measured for each BR–RLC (Fig. 2). When a fluorescence dipole becomes more parallel to the muscle fibre axis, $P_{||}$ increases and $^xP_{\perp}$ and $^yP_{\perp}$ decrease. P ratios varied systematically between the four experimental conditions studied (Fig. 2), indicating different orientations of the RLC region of myosin heads with respect to the fibre axis. P ratios also depended on which BR–RLC mutant was exchanged into fibres, as expected from the different orientations of their BR dipoles in the RLC coordinate frame (Fig. 1b). The orientation of each BR dipole was determined from these P ratios, taking account of fluorescence

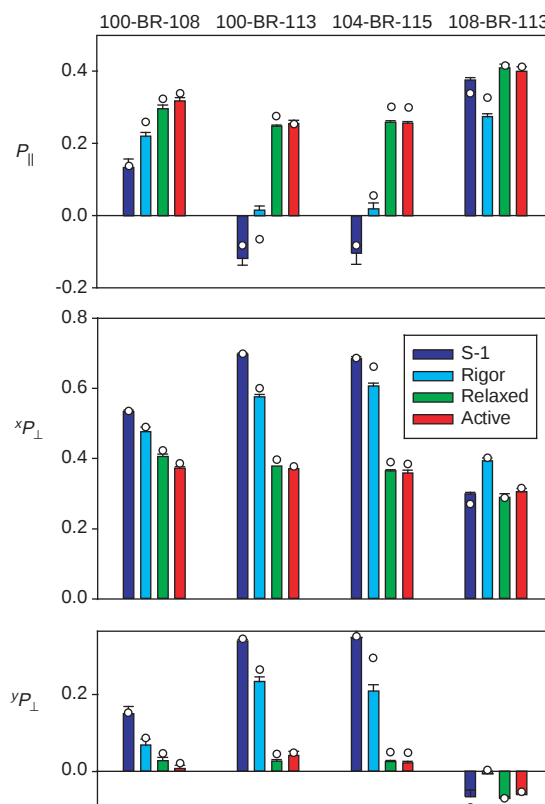


Figure 2 Polarization ratios $P_{||}$, $^xP_{\perp}$ and $^yP_{\perp}$ from single muscle fibres containing each of the four BR–RLCs. Relaxed, active and rigor values are from BR–RLC exchanged into endogenous myosin heads in 6–9 fibres for each BR–RLC at sarcomere length 2.4 μ m. S-1 values are for exogenous S-1 containing BR–RLC, bound to actin filaments in 4–7 fibres at sarcomere length 4.0 μ m in the absence of ATP. The circles were calculated from the best-fit gaussian model parameters given in Table 1. Bars are means $\pm$ s.e. Temperature, 10 °C.

depolarization caused by independent probe wobble on the nano-second timescale of the fluorescence decay^{16,28}.

The three-dimensional orientation of the RLC region can be deduced from the orientations of the four bifunctional rhodamine dipoles, knowing the crystallographic coordinates of the residues where each rhodamine probe was attached (Fig. 1b). The orientation of the RLC region is described using a coordinate frame defined by two α -helices; the ‘lever’ axis is parallel to the long heavy-chain helix that forms the backbone of the putative lever arm of the myosin motor, and the ‘hook’ axis is parallel to the short heavy-chain helix next to the myosin head–rod junction. The orientation of the RLC frame with respect to the fibre axis is expressed as the tilt angle (β) between the lever and fibre axes, and the twist angle (γ) that describes rotation of the RLC region around the lever axis (Fig. 3, inset). β can take values between 0 and 180°; an increase in β for a myosin head attached to actin would cause filament sliding in the direction corresponding to muscle shortening. γ can take values between –180° and +180°; $\gamma = 0^\circ$ when the hook axis is coplanar with the lever and fibre axes and its carboxy terminus points towards the M-line of the sarcomere. Changes in γ could also produce displacement of the head–rod junction, leading to filament sliding.

The RLC region is not expected to have the same orientation (β , γ) in every myosin head in a muscle fibre. The two heads in each myosin molecule and each intermediate in the ATPase cycle, including populations attached to or detached from actin, may have different orientations. To accommodate a distribution of orientations, polarized fluorescence data from the BR–RLCs were

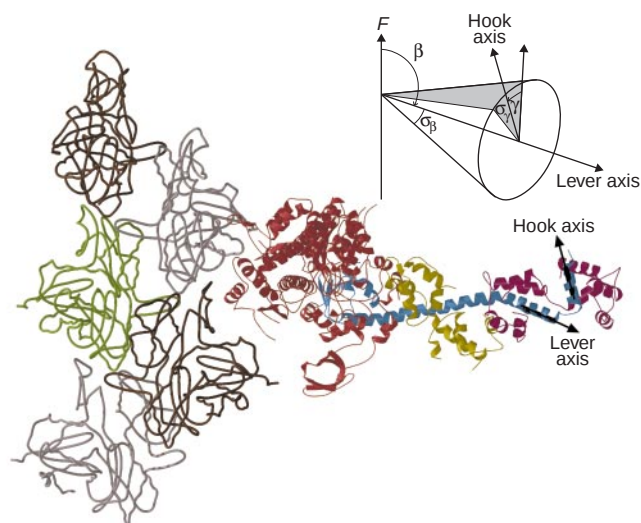


Figure 3 Conformation of the myosin head (catalytic domain, red; heavy-chain residues 707–843, blue; regulatory and essential light chains, magenta and yellow, respectively) bound in rigor to the actin filament (brown, grey, green). Inset: the orientations of the lever and hook axes with respect to the filament axis (F), corresponding to the peak tilt and twist angles (β_0 and γ_0 , respectively) in Table 1, rigor; the cone and its shaded segment show the range of tilt and twist dispersions ($\pm\sigma_\beta$ and $\pm\sigma_\gamma$). Molecular graphics prepared using Molscrip⁴³ and Raster3D⁴⁴.

fitted with gaussian distributions of β and γ with four independent parameters: peak angles β_0 and γ_0 , and corresponding dispersions σ_β and σ_γ . After removing the effects of rapid probe wobble, two order parameters, $\langle P_2 \rangle$ and $\langle P_4 \rangle$, were obtained from the steady-state polarized fluorescence measurements for each BR–RLC²⁸, giving eight observed quantities, twice the number of parameters in the two-dimensional gaussian distribution.

The gaussian orientation parameters that best fitted the order parameters derived from fluorescence polarization ratios $P_{||}$, $^*P_{\perp}$ and $^*P_{\perp}$ are given in Table 1. Polarization ratios calculated from these gaussian parameters (Fig. 2, circles) were, with few exceptions, in good agreement with the observed values (Fig. 2, bars). The peak tilt angle, β_0 , for exogenous S-1 was 111°, which is identical to the value estimated by fitting the crystal structure of skeletal S-1 (ref. 7) into density maps calculated from electron micrographs of isolated actin filaments decorated with S-1 fragments⁹. However, the peak twist angle, γ_0 , was 23°, which is significantly different from the corresponding angle, –9°, inferred from the electron-microscopic density maps. RLC was missing from the S-1 fragments used for those studies, so the orientation of the catalytic domain of the myosin head was probably determined more accurately than that of its light-chain domain. As the angle between these two domains can vary by up to 70° (refs 7, 8), the most likely interpretation of our data is that the RLC region has twisted by about 30° with respect to the catalytic domain, which is bound to actin in the conformation determined by electron microscopy⁹.

The best-fit values of β_0 , σ_β , γ_0 and σ_γ for endogenous myosin heads in rigor were almost identical to those obtained with exogenous S-1 (Table 1). In the absence of ATP, endogenous myosin heads therefore bind to actin with their RLC region in almost the same conformation as in the untethered actin–myosin-head complex. We calculated an atomic model for endogenous myosin heads bound to actin in rigor (Fig. 3) from the published model for exogenous S-1 fragments bound to actin^{7,9} by rotating the light-chain domain with respect to the catalytic domain to match the best-fit values of β_0 and γ_0 . The hook helix points towards the top of Fig. 3 (corresponding to the M-line in the sarcomere), the

Table 1 Tilt and twist of the RLC region

	β_0	σ_β	γ_0	σ_γ
S-1	111° {109, 113°}	21° {19, 23°}	23° {22, 24°}	25° {24, 27°}
S-1 (ref. 9)	111°		–9°	
Rigor	108° {105, 111°}	24° {20, 28°}	29° {26, 32°}	29° {27, 31°}
Relaxed	100° {97, 104°}	37° {34, 39°}	9° {3, 17°}	52° {50, 54°}
Active isometric	96° {91, 120°}	37° {11, 43°}	12° {2, 47°}	53° {43, 57°}

Best-fit gaussian parameters for the two-dimensional orientation distribution of the RLC region: tilt distribution, $\exp[-(\beta - \beta_0)^2/2\sigma_\beta^2]$, with peak β_0 and dispersion σ_β ; twist distribution, $\exp[-(\gamma - \gamma_0)^2/2\sigma_\gamma^2]$, with peak γ_0 and dispersion σ_γ , $\beta > 90^\circ$ when the RLC end of the myosin head is tilted towards the Z-line of the sarcomere (downwards in Fig. 3). An increase in γ_0 denotes a counter-clockwise rotation of the RLC around the lever axis as viewed from the head–rod junction. Values in brackets are 95% confidence limits estimated as described in Methods. Reduced χ^2 values (with four degrees of freedom) for the best-fit gaussian parameters were: 0.42 (S-1), 5.17 (rigor), 7.69 (relaxed) and 1.52 (active).

correct orientation for pulling the myosin filament downwards during muscle shortening. Orientational dispersion of the RLC region is represented as a cone centred on the lever axis (Fig. 3, inset); the cone width is $2\sigma_\beta$ in the gaussian model, and the shaded segment represents $2\sigma_\gamma$. In these conditions σ_γ happens to have the same value as γ_0 , so the right-hand edge of the shaded segment corresponds to the $\gamma = 0$ axis. With a 10-nm lever arm, the tilt dispersion $2\sigma_\beta$ corresponds to an 8-nm range of displacement of the head–rod junction along the filament axis. As actin sites are spaced 5.5 nm apart on each helical strand of the filament, this range would accommodate the mismatch in periodicities between the actin and myosin filaments.

β_0 was slightly smaller in relaxed fibres than in rigor (Table 1). γ_0 was also smaller in relaxed fibres; the hook helix points more directly towards the M-line. Both σ_β and σ_γ were larger in relaxation. The values of all four gaussian parameters in active contraction were similar to those in relaxation. These conclusions broadly confirm the interpretation of previous studies with monofunctional RLC probes of unknown orientation in the RLC frame^{16–18,29}. The relatively small differences between RLC orientation distributions in rigor, relaxation and isometric contraction provide further evidence that force generation is associated with motions of only a small fraction of the myosin heads^{25–27}. We estimated this fraction and measured the conformational change in the force-generating heads by applying rapid length steps to actively contracting muscle fibres.

Motions of force-generating heads

When the length of an actively contracting muscle fibre is abruptly decreased, the imposed sliding between actin and myosin filaments leads to an elastic decrease in force, followed by a very rapid force recovery that has been proposed to be due to the elementary force-generating transition in the actin-attached myosin heads³⁰. We measured the changes in orientation of the RLC region initiated by a shortening step of 3.6 nm per half-sarcomere with 50- μ s time resolution (Fig. 4a) by modulating the polarization of the excitation light^{16,31}. This method measures Q polarization ratios (see Methods), which are equal to the corresponding P ratios for rhodamines²⁹, because the absorption and emission dipoles are colinear²². With the gaussian orientation model, the Q transients indicate a small increase in β_0 , by 2.3°, and a small decrease in γ_0 , by 1.7° (Fig. 4b). β_0 increases during the length step itself, showing that the myosin head is elastic^{31,32}, and during rapid-force recovery, as expected for a motor mechanism that involves both active tilting and elastic bending of the myosin head^{19,33}.

The gaussian orientation model used in Fig. 4b implies that all myosin heads in the fibre respond to the length step. However, we have argued above that only a small fraction of the heads may bear

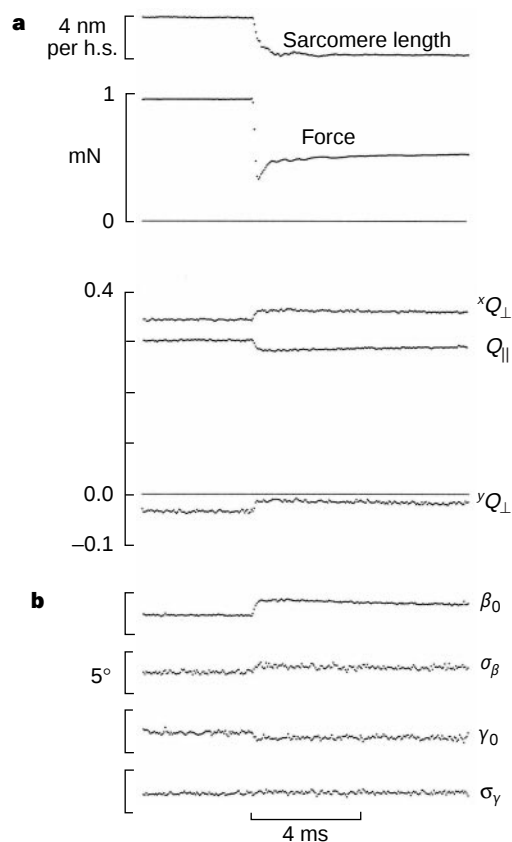


Figure 4 Mechanical and RLC-region orientation transients. **a**, Sarcomere length, force, and $Q_{\parallel}$, $Q_{\perp}$ and Q_{γ} transients from a fibre containing 100-BR-108 RLC, produced by a 150-μs shortening step of amplitude 3.6 nm per half sarcomere (h.s.) during active contraction; initial sarcomere length, 2.4 μm; temperature, 12 °C. Q traces averaged from 24 repeats of the length-step protocol. **b**, Gaussian model parameters obtained by fitting Q transients from fibres containing 100-BR-108, 104-BR-115 and 108-BR-113 RLCs (one fibre for each BR-RLC). For these three fibres the pre-step values of β_0 , σ_{β} , γ_0 and σ_{γ} were 97.3°, 21.5°, 41.3° and 40.1°, respectively.

the force of isometric contraction. We therefore re-analysed the Q transients, assuming that they arise from two populations of myosin heads: a static population that does not respond to the length step, and a moving population in which the myosin head-rod junction is displaced axially by 3.6 nm, the imposed filament sliding. In the moving population the light-chain region was assumed to tilt and twist around a pivot at the SH-1 helix (Cys 707), as in Fig. 3. This two-population analysis showed that the fraction of heads responding to the length step was 12%, the same as the value estimated previously with monofunctional RLC probes from the nonlinear dependence of Q transients on length-step amplitude¹⁶. The peak tilt angle (β_0) for the moving heads increased by 19°, and the peak twist angle (γ_0) decreased by 13°, as shown in Fig. 5a. These angle changes relate to a particular length-step amplitude (3.6 nm) rather than the total range of the working stroke. Isolation of elastic and working-stroke contributions to the motion will require further experiments using different length-step amplitudes¹⁶.

Both the increase in tilt angle (β) and the decrease in twist angle (γ) drive filament sliding. The axial displacement of the myosin head-rod junction was calculated from these angles under the previous assumption that the catalytic domain binds to actin in a fixed conformation (Fig. 5b). The crystallographic conformation of skeletal S-1 in the absence of ATP (circle) may represent the end of

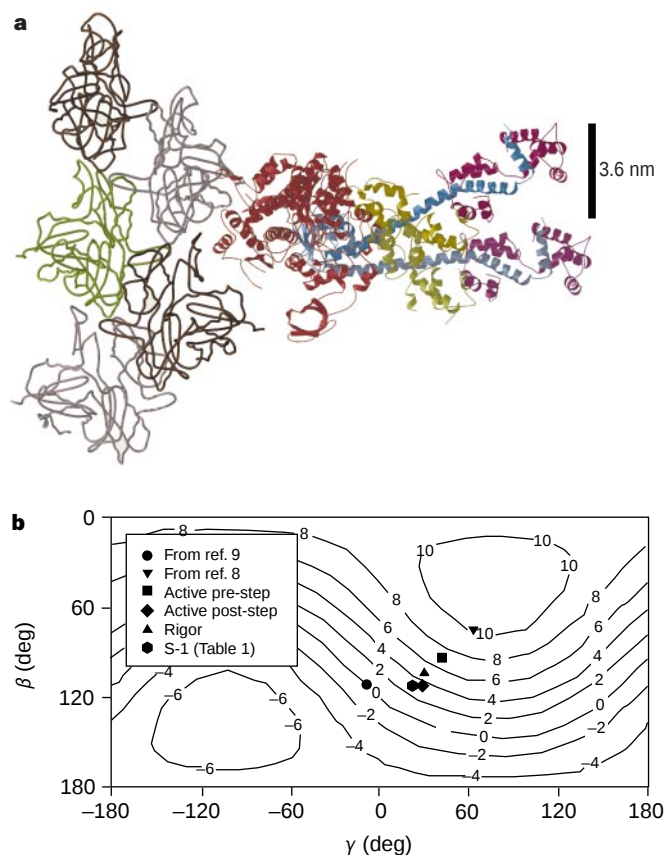


Figure 5 Tilt and twist of the RLC region. **a**, Change in conformation of the fraction of myosin heads that respond to filament sliding produced by imposing a shortening step of 3.6 nm per half-sarcomere during active contraction. Colour scheme as Fig. 3; lighter shades represent the post-step conformation. **b**, Contour plot of the axial displacement of the myosin head-rod junction (nm) relative to that in ref. 9, with superimposed symbols indicating the conformations of the RLC region described in the text. β and γ for the ADP·AlF₄ conformation were calculated by modelling the RLC onto the published structure⁸ assuming that the essential-light-chain/RLC interface was as in skeletal myosin⁷ and that the catalytic domain is fixed on actin.

the working stroke^{7,9}; that of a smooth muscle myosin fragment in the presence of ADP·AlF₄ (inverted triangle) may represent the start of the working stroke⁸. These two conformations differ by 36° and 75° in β and γ respectively, producing a 10-nm axial displacement. The orientations of the RLC region of the force-generating population of heads in active muscle (pre-step, square; post-step, diamond) lie between the two crystallographic conformations, in the region where axial displacement is most sensitive to β and γ . The post-step active conformation is close to the rigor conformation in the same set of fibres (triangle), and to that for exogenous S-1 (hexagon; from Table 1). Thus, our results show that the motions of the RLC region predicted by the crystallographic models do take place in working muscle fibres, and that filament sliding involves both tilting and twisting of this domain.

Discussion

Using the myosin regulatory light chain as a model system, we have shown that the bifunctional rhodamine fluorescence-polarization method can measure the orientation of a protein domain in a cellular environment on the sub-millisecond timescale. The method requires that the static structure of the protein is known, and that suitable double-cysteine mutants of the protein can be expressed, labelled specifically with bifunctional rhodamine and re-introduced

into a functional complex. The resulting information about real-time motions of protein domains in such complexes should complement high-resolution structures obtained by X-ray crystallography and nuclear magnetic resonance.

The structural information obtained by this approach depends on the number and relative orientation of the cysteine pairs used, and on the symmetry and orientational order of the macromolecular system under investigation. In systems with cylindrical symmetry, like a muscle fibre, both the tilt and twist of a well-ordered domain can be determined with respect to the symmetry axis of the system, but azimuthal motions around this axis cannot be detected. If there is substantial orientational disorder, or if more than one population of oriented domains is present, it is more difficult to characterize the distribution of tilt and twist. These limitations could be overcome by measurements on protein ensembles with lower symmetry or on single molecules^{34–36}. □

Methods

RLCs labelled with bifunctional rhodamine were obtained in 10% overall yield and 90% purity by reacting mutants of chicken gizzard RLC, expressed in *Escherichia coli*²⁷, with BR-I₂. The 1:1 two-site labelling was confirmed for all labelled RLCs by electrospray mass spectroscopy, and gave two diastereomeric labelled proteins because of restricted rotation of the carboxy-substituted phenyl ring about the single bond joining it to the three coplanar rings that comprise the fluorophore³⁸. In 104–BR–115 and 108–BR–113 RLCs the diastereoisomers were separable. Steady-state experiments with each isomer of 108–BR–113 gave identical polarization ratios. In all other cases data were obtained from mixed diastereoisomers.

Experimental methods and solutions used for incorporating labelled RLCs into rabbit psoas muscle fibres, and for measuring steady-state fluorescence polarization ratios, $P_{\parallel}$, $^xP_{\perp}$ and $^yP_{\perp}$, have been described^{16,29,39}. The RLC content of the fibres was measured using silver-stained SDS gels, with the essential light-chain concentration as internal control²¹. Between 40 and 90% of the native RLC was extracted from the fibre and replaced by the same amount of BR–RLC, so the total RLC content of the fibre remained constant, within the precision of the measurements. BR–RLC exchange had no significant effect on active force production or relaxation. $P_{\parallel} = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + I_{\perp})$; $^xP_{\perp} = (I_{\perp}^x - I_{\perp}^y) / (I_{\perp}^x + I_{\perp}^y)$; $^yP_{\perp} = (I_{\perp}^y - I_{\perp}^x) / (I_{\perp}^y + I_{\perp}^x)$; I is the component of fluorescence intensity polarized relative to the fibre axis as denoted by the post-subscript, with excitation polarized as denoted by the pre-subscript; superscripts x and y indicate that the propagation axis of observed fluorescence was at 0° and 90°, respectively, relative to that of excitation, in the plane perpendicular to the fibre axis. Polarization ratios $Q_{\parallel}$, $^xQ_{\perp}$ and $^yQ_{\perp}$ were measured at 50-μs time resolution using a photoelastic modulator in the excitation optics¹⁶; $Q_{\parallel} = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + I_{\perp})$; $^xQ_{\perp} = (I_{\perp}^x - I_{\perp}^y) / (I_{\perp}^x + I_{\perp}^y)$; $^yQ_{\perp} = (I_{\perp}^y - I_{\perp}^x) / (I_{\perp}^y + I_{\perp}^x)$. P or Q ratios were interpreted in terms of the second- and fourth-rank order parameters ($\langle P_2 \rangle$ and $\langle P_4 \rangle$) of the orientational distribution of the probe transition dipole moment with respect to the fibre axis, taking into account effects of rapid probe motions that are fast compared with the decay of the excited state²⁸. For each BR–RLC probe, in all conditions studied, this motion was equivalent to wobble in a cone of semi-angle 20–30°, similar to the 18–32° range observed with twelve monofunctional rhodamine probes on the RLC^{17,29}. Mobility of the bifunctionally attached probes is probably due to flexibility in the linker arms of bifunctional rhodamine (Fig. 1a), allowing some independent motion of the rhodamine fluorophore with respect to its RLC attachments. Assuming that this motion is cylindrically symmetrical, the best available estimate of average orientation of the bifunctional rhodamine dipole is the line joining the β -carbons of the two residues that were replaced by cysteines, as location of these atoms would not be altered by mutagenesis in the absence of changes in backbone conformation. The β -carbon pairs were separated by 1.0–1.6 nm in the four BR–RLCs, whereas the linker arms could potentially span distances in the range 0.4–2.4 nm.

We calculated predicted values of the order parameters ($\langle P_2 \rangle$ and $\langle P_4 \rangle$) for each BR–RLC for a gaussian orientational distribution of the RLC region with parameters β_0 , σ_{β} , γ_0 and σ_{γ} as defined in Table 1, using the orientations of each bifunctional rhodamine dipole in the RLC coordinate frame and the comple-

tion theorem. To present a more direct representation of the β distribution of the RLC region, we did not include a $\sin\beta$ weighting factor, in contrast with models used previously for axial orientation distributions of monofunctional probe dipoles^{16,17,29,31,39}. Including the $\sin\beta$ weighting factor did not affect our conclusions. We defined the RLC frame (Fig. 1b) in terms of the 'lever' axis, joining α -carbons of Phe 814 and Asn 825 of the long heavy-chain helix near the RLC, and the hook axis, joining the midpoints of α -carbons of the pairs Phe 836/Ile 838 and Met 832/Leu 834. Assuming that the bifunctional rhodamine dipole is aligned with β -carbons of the labelled residues, with atomic coordinates for chicken skeletal myosin provided by I. Rayment⁷, the 100–BR–108, 100–BR–113, 104–BR–115 and 108–BR–113 dipoles have angular coordinates: $\theta = 50^\circ, 53^\circ, 90^\circ$ and 81° ; $\phi = 132^\circ, 76^\circ, 66^\circ$ and 30° , respectively, in this frame, where θ is the angle between the dipole and lever axes and ϕ is the angle between two planes: one defined by the lever and dipole axes and the other by the lever and hook axes. Increasing ϕ represents counter-clockwise rotation of the dipole around the lever axis as viewed from the myosin head–rod junction. The four gaussian model parameters were adjusted using a weighted Marquardt–Levenberg algorithm⁴⁰ to minimize deviations between the predicted and measured values of $\langle P_2 \rangle$ and $\langle P_4 \rangle$ for each BR–RLC. Low-resolution global searches of the parameter space showed that these best-fit parameters were likely to be global minima. We used Monte-Carlo simulations around each solution to estimate 95% confidence limits for each of the four fitted parameters, corresponding to values of χ^2 that were 4.0 larger than the minimum value⁴⁰. The effect of systematic errors in the orientations (θ , ϕ) of one of the four bifunctional rhodamine dipoles in the RLC coordinate frame was assessed by repeating the fitting procedure after changing one of the eight values of θ or ϕ by $\pm 10^\circ$; the resulting average change in the best-fit values of β_0 and γ_0 was $\sim 2^\circ$. As the BR–RLCs bind to the myosin heavy chain in a conformation close to the native one, these results indicate that errors in gaussian fit parameters resulting from systematic errors in θ and ϕ are likely to be negligible. Because of the symmetry of the muscle sarcomere, the orientations (β , γ) and ($180^\circ - \beta$, $180^\circ + \gamma$) cannot be distinguished by fluorescence polarization measurements alone. The solutions described in the text and Table 1 are justified by three criteria: they are consistent with the conformation of exogenous S-1 bound to isolated actin filaments⁹; the hook helix points towards rather than away from the M-line (Fig. 3); and, with these solutions, length-step experiments in both active contraction (Fig. 4) and rigor (data not presented) showed a tilt of the RLC region in the expected direction, towards the Z-line, for a shortening step.

We recorded fluorescence polarization transients and sarcomere length following rapid length steps, as described¹⁶. After peak isometric force had been reached in an activation, 24 repeated cycles of shortening followed by re-stretching to the original length were applied; at least 200 ms was allowed for recovery between the re-stretch and the next shortening. Force, motor position and $Q_{\parallel}$, $^xQ_{\perp}$ and $^yQ_{\perp}$ signals were averaged from the 24 cycles. We used the same protocol in separate activations for recording striation spacing and Q signals with x - and y -illumination. In some experiments (Fig. 4a) the sarcomeres continued to shorten during the force recovery, probably accompanied by stretching of weaker regions at the fibre ends. The delayed component of Q transients from RLC probes is still present in the absence of such delayed shortening^{16,31}. The values of $Q_{\parallel}$, $^xQ_{\perp}$ and $^yQ_{\perp}$ before each length step were somewhat different from the corresponding P ratios in Fig. 2, probably because of systematic changes in Q ratios produced by the many shortening–stretch cycles and repeated activations that were averaged to obtain an acceptable signal:noise ratio at this time-resolution¹⁶. Similarly, the values of β_0 , σ_{β} , γ_0 and σ_{γ} in the length-step experiments differ from those for active isometric contraction in Table 1. In 20 fibres (7 with 100–BR–108, 5 with 104–BR–115, and 8 with 108–BR–113) the values of β_0 , σ_{β} , γ_0 and σ_{γ} were: before the length step, 93.3°, 22.1°, 31.2°, 43.8°; after the length step, 95.6°, 22.9°, 29.7°, 43.5°, respectively.

BR–RLCs were exchanged into rabbit myosin subfragment-1 (S-1) prepared from dorsal and hind-leg white muscle³⁹. The BR–RLC-labelled S-1 was repurified by gel filtration. Single psoas muscle fibres were slowly stretched to 4 μm sarcomere length in relaxing solution, transferred to EDTA rigor solution, then incubated in $\sim 6 \text{ mg ml}^{-1}$ BR–RLC-labelled S-1 at 10°C for 60 min³⁹. Polarized fluorescence intensities were measured after two washes in rigor solution, and again after washing in relaxing solution, when 10–20%

fluorescence remained. This residual fluorescence was subtracted before calculating P ratios. BR–RLCs were exchanged into scallop myofibrils as described⁴¹, and the steady-state Mg^{2+} –ATPase activity of the resulting hybrids, containing a 1:1 ratio of BR–RLC to native scallop essential light chain, was measured using a pH-Stat⁴⁷.

Received 27 April; accepted 29 June 1999.

- Corrie, J. E. T., Craik, J. S. & Munasinghe, V. R. N. A homobifunctional rhodamine for labeling proteins with defined orientations of a fluorophore. *Bioconjug. Chem.* **9**, 160–167 (1998).
- Reedy, M. K., Holmes, K. C. & Tregear, R. T. Induced changes in orientation of the cross-bridges of glycerinated insect flight muscle. *Nature* **207**, 1276–1280 (1965).
- Huxley, H. E. The mechanism of muscle contraction. *Science* **164**, 1356–1366 (1969).
- Huxley, A. F. Muscular contraction (review lecture) *J. Physiol. (Lond.)* **243**, 1–43 (1974).
- Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. Atomic structure of the actin: DNase I complex. *Nature* **347**, 37–43 (1990).
- Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. Atomic model of the actin filament. *Nature* **347**, 44–49 (1990).
- Rayment, I. et al. Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science* **261**, 50–58 (1993).
- Dominguez, R., Freyzon, Y., Trybus, K. M. & Cohen, C. Crystal structure of a vertebrate smooth muscle myosin motor domain and its complex with the essential light chain: visualization of the pre-power stroke state. *Cell* **94**, 559–571 (1998).
- Rayment, I. et al. Structure of the actin–myosin complex and its implications for muscle contraction. *Science* **261**, 58–65 (1993).
- Whittaker, M. et al. A 35-Å movement of smooth muscle myosin on ADP release. *Nature* **378**, 748–753 (1995).
- Holmes, K. C. The swinging lever arm hypothesis of muscle contraction. *Curr. Biol.* **7**, R112–R118 (1997).
- Cooke, R. Actomyosin interaction in striated muscle. *Physiol. Rev.* **77**, 671–697 (1997).
- Goldman, Y. E. Wag the tail: structural dynamics of actomyosin. *Cell* **93**, 1–4 (1998).
- Uyeda, T. Q. P., Abramson, P. D. & Spudich, J. A. The neck region of the myosin motor domain acts as a lever arm to generate movement. *Proc. Natl Acad. Sci. USA* **93**, 4459–4464 (1996).
- Anson, M., Geeves, M. A., Kurzawa, S. E. & Manstein, D. J. Myosin motors with artificial lever arms. *EMBO J.* **15**, 6069–6074 (1996).
- Hopkins, S. C., Sabido-David, C., Corrie, J. E. T., Irving, M. & Goldman, Y. E. Fluorescence polarization transients from rhodamine isomers on the myosin regulatory light chain in skeletal muscle fibers. *Biophys. J.* **74**, 3093–3110 (1998).
- Sabido-David, C. et al. Orientation changes of fluorescent probes at five sites on the myosin regulatory light chain during contraction of single skeletal muscle fibres. *J. Mol. Biol.* **279**, 387–402 (1998).
- Baker, J. E., Brust-Mascher, I., Ramachandran, S., LaConte, L. E. W. & Thomas, D. D. A large and distinct rotation of the myosin light chain domain occurs upon muscle contraction. *Proc. Natl Acad. Sci. USA* **95**, 2944–2949 (1998).
- Dobbie, I. et al. Elastic bending and active tilting of myosin heads during muscle contraction. *Nature* **396**, 383–387 (1998).
- Simmons, R. M. & Szent-Györgyi, A. G. Control of tension development in scallop muscle fibres with foreign regulatory light chains. *Nature* **286**, 626–628 (1980).
- Moss, R. L., Giulian, G. G. & Greaser, M. L. Physiological effects accompanying removal of myosin LC₂ from skinned skeletal muscle fibers. *J. Biol. Chem.* **257**, 8588–8591 (1982).
- Penzkofer, A. & Wiedmann, J. Orientation of transition dipole moments of rhodamine 6G determined by excited state absorption. *Opt. Commun.* **35**, 81–86 (1980).
- Xie, X. et al. Structure of the regulatory domain of scallop myosin at 2.8 Å resolution. *Nature* **368**, 306–312 (1994).
- Jansco, A. & Szent-Györgyi, A. G. Regulation of scallop myosin by the regulatory light chain depends on a single glycine residue. *Proc. Natl Acad. Sci. USA* **91**, 8762–8766 (1994).
- Cooke, R., Crowder, M. S. & Thomas, D. D. Orientation of spin labels attached to cross-bridges in contracting muscle fibers. *Nature* **300**, 776–778 (1982).
- Huxley, H. E. & Kress, M. Crossbridge behaviour during muscle contraction. *J. Muscle Res. Cell Motil.* **6**, 153–161 (1985).
- Linari, M. et al. The stiffness of skeletal muscle in isometric contraction and rigor: the fraction of myosin heads bound to actin. *Biophys. J.* **74**, 2459–2473 (1998).
- Dale, R. E. et al. Model-independent analysis of mobility and orientation of fluorescent probes in muscle fibers. *Biophys. J.* **76**, 1606–1618 (1999).
- Sabido-David, C. et al. Steady-state fluorescence polarization studies of the orientation of myosin regulatory light chains in single skeletal muscle fibers using pure isomers of iodoacetamidotetramethylrhodamine. *Biophys. J.* **74**, 3083–3092 (1998).
- Huxley, A. F. & Simmons, R. M. Proposed mechanism of force generation in striated muscle. *Nature* **233**, 533–538 (1971).
- Irving, M. et al. Tilting of the light-chain region of myosin during step length changes and active force generation in skeletal muscle. *Nature* **375**, 688–691 (1995).
- Lombardi, V. et al. Elastic distortion of myosin heads and repriming of the working stroke in muscle. *Nature* **374**, 553–555 (1995).
- Irving, M., Lombardi, V., Piazzesi, G. & Ferenczi, M. A. Myosin head movements are synchronous with the elementary force-generating process in muscle. *Nature* **357**, 156–158 (1992).
- Funatsu, T., Harada, Y., Tokunaga, M., Saito, K. & Yanagida, T. Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature* **374**, 555–559 (1995).
- Ishijima, A. et al. Simultaneous observation of individual ATPase and mechanical events by a single myosin molecule during interaction with actin. *Cell* **92**, 161–171 (1998).
- Warshaw, D. M. et al. Myosin conformational states determined by single fluorophore polarization. *Proc. Natl Acad. Sci. USA* **95**, 8034–8039 (1998).
- Rowe, T. & Kendrick-Jones, J. Chimeric myosin regulatory light chains identify the subdomain responsible for regulatory function. *EMBO J.* **11**, 4715–4722 (1992).
- Oki, M. Recent advances in atropisomerism. *Top. Stereochem.* **14**, 1–79 (1983).
- Ling, N., Shrimpton, C., Sleep, J., Kendrick-Jones, J. & Irving, M. Fluorescent probes of the orientation of myosin regulatory light chains in relaxed, rigor, and contracting muscle. *Biophys. J.* **70**, 1836–1846 (1996).
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in Fortran: The Art of Scientific Computing* 2nd edn (Cambridge Univ. Press, Cambridge, UK, 1992).
- Chantler, P. D. & Szent-Györgyi, A. G. Regulatory light-chains and scallop myosin: full dissociation, reversibility and co-operative effects. *J. Mol. Biol.* **138**, 473–492 (1980).
- Esnouf, R. M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 133–138 (1997).
- Kraulis, P. J. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).

Acknowledgements. This work was supported by the UK MRC, Wellcome Trust, NIH, MDA and AHA. We thank S. Howell for mass spectrometric analyses, M. Hirschberg and I. Dobbie for help with molecular graphics, and V. R. N. Munasinghe for technical support with BR-1.

Correspondence and requests for materials should be addressed to M.I. (e-mail: malcolm.irving@kcl.ac.uk).

Discovery of the acoustic Faraday effect in superfluid $^3\text{He-B}$

Y. Lee, T. M. Haard, W. P. Halperin & J. A. Sauls

Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA

Acoustic waves provide a powerful tool for studying the structure of matter. For example, the speed, attenuation and dispersion of acoustic waves can give useful information on molecular forces and the microscopic mechanisms of absorption and scattering of acoustic energy. In solids, both compression and shear waves occur—longitudinal and transverse sound, respectively. But normal liquids do not support shear forces and consequently transverse waves do not propagate in liquids, with one notable exception. In 1957 Landau predicted¹ that the quantum-liquid phase of helium-3 might support transverse sound waves at sufficiently low temperatures, the restoring forces for shear waves being supplied by the collective quantum behaviour of the particles in the fluid. Such shear waves will involve displacements of the fluid transverse to the direction of propagation, and so define a polarization direction similar to that of electromagnetic waves. Here we confirm experimentally the existence of transverse sound waves in superfluid $^3\text{He-B}$ by observing the rotation of the polarization of these waves in the presence of a magnetic field. This phenomenon is the acoustic analogue of the magneto-optic Faraday effect, whereby the polarization direction of an electromagnetic wave is rotated by a magnetic field applied along the propagation direction.

Superfluidity in ^3He results from the binding of the ^3He particles with nuclear spin $s = 1/2$ into molecules called “Cooper pairs” with binding energy 2Δ (refs 2–5). The pairs undergo a type of Bose–Einstein condensation, which has a close analogy to the Bardeen–Cooper–Schrieffer condensation⁶ phenomenon associated with superconductivity in metals. One important difference is that the pairs that form the condensate in ^3He have total spin $S = 1$ and an orbital wavefunction with relative angular momentum $L = 1$ (*p-wave*). This is in contrast to superconductors which are formed with Cooper pairs of electrons having $S = 0$ and $L = 0$ (*s-wave*) or, as is the case of high-temperature superconductors, $S = 0$ and $L = 2$ (*d-wave*). In superfluid ^3He , the spin and orbital angular momentum vectors are locked at a fixed angle to one another. This is called broken relative spin–orbit symmetry^{2,3}. The equilibrium superfluid state is described as a condensate of Cooper pairs with a total angular momentum $J = L + S = 0$. In addition, the Cooper pairs can be resonantly excited by sound waves to quantum states with total angular momentum $J = 2$ (ref. 7). This is reminiscent of diatomic molecules which have similar excited states.

The above description applies to the B-phase of superfluid ^3He , the most stable phase at low pressure. The acoustic Faraday effect occurs in $^3\text{He-B}$ as a consequence of spontaneously broken relative spin–orbit symmetry⁸. An applied field magnetically polarizes the spins of the Cooper pairs which, through coupling to their orbital motion, rotates the polarization of transverse sound. The rotational excitations of Cooper pairs are essential⁸ to our observation of the propagation of transverse acoustic waves in $^3\text{He-B}$ because they significantly increase the sound velocity making the sound mode much easier to detect; the closer the sound energy is to the energy of the Cooper-pair excited state, the stronger is this effect. Furthermore, the Cooper-pair excited states have a linear Zeeman splitting with magnetic field^{9,10}. Of the five $(2J + 1)$ Zeeman substates there is one, $m_J = +1$, which couples to right circularly polarized transverse

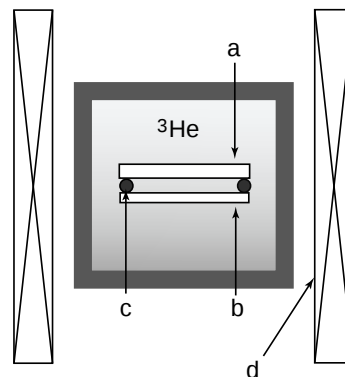


Figure 1 Acoustic cavity for transverse sound. Our short-path-length acoustic cavity consists of two quartz transducers 6.3 mm in diameter, labelled a and b. Their ground planes are face-to-face, and the spacing between the transducers is established by two parallel gold-coated tungsten wires (c). One of the cavity walls is a 12-MHz, AC-cut transducer (b) used for transverse sound excitation and detection. The other face of the cavity is a 17-MHz, X-cut transducer (a) for excitation and detection of longitudinal sound, which we used to determine the cavity length of 31 μm . The acoustic cavity is immersed in liquid ^3He at a pressure near 4.5 bar. The experimental cell is cooled by a copper nuclear demagnetization refrigerator (not shown). A superconducting solenoid (d) is placed outside the sample liquid enclosed within a superconducting shield. Temperatures were determined by SQUID-based paramagnetic salt thermometry.

sound, and a second, $m_J = -1$, couples to left circularly polarized sound. Thus the speeds of these two transverse waves are different in a magnetic field. We call this property acoustic birefringence. It leads to the acoustic Faraday effect where the magnetic field rotates the polarization direction of linearly polarized sound. Our measurements show that the rotation angle can be as large as 1.4×10^7 degrees $\text{cm}^{-1} \text{T}^{-1}$, much larger than the usual magneto-optical Faraday effect¹¹.

Excitation and detection of transverse sound is provided by a high-Q ($\sim 3,000$), AC-cut, quartz transducer with a fundamental resonance frequency of 12 MHz. It generates and detects shear waves with a specific linear polarization. The detection method is based on measurement of the electrical impedance of the transducer using a frequency-modulated spectrometer¹². All measurements were performed at 82.26 MHz, the seventh harmonic of the transducer, with frequency modulation at 400 Hz and a modulation amplitude of 3 kHz. The electrical impedance of the transducer is a direct measure of the acoustic impedance of the surrounding liquid ^3He in the acoustic cavity that is shown in Fig. 1. Linearly polarized waves are excited by the transducer, reflected from the opposite surface of the acoustic cell, and detected by the same transducer. Under conditions of high attenuation there is no reflected wave, and the acoustic response is determined by the bulk acoustic impedance, $Z_a = \rho\omega/q$ where ρ is the density of the liquid, ω is the sound frequency, $q = k + i\alpha$ is the complex wavenumber, α is the attenuation, and $2\pi/k$ is the wavelength. A change in either the attenuation or the phase velocity, $C_\phi = \omega/k$, produces a change in the impedance, Z_a .

On cooling into the highly attenuating superfluid region, the acoustic response varies smoothly with temperature (Fig. 2). If the attenuation is low, there is interference between the source and reflected waves which modulates the local acoustic impedance detected by the transducer. Consequently, the acoustic response oscillates as the phase velocity changes with temperature. The oscillations in Fig. 2 at low temperatures correspond to interference between outgoing and reflected waves, and so they indicate the existence of some form of propagating wave. Each period of the oscillations corresponds to a change in velocity sufficient to increase, or decrease, by unity the number of half wavelengths in the cavity. The amplitude of the oscillations increases as the

temperature is reduced, indicating that attenuation of the sound mode decreases with decreasing temperature.

The features labelled A and B in Fig. 2 are identified with known physical processes for sound absorption in superfluid $^3\text{He-B}$ (refs 3, 13, 14). Feature A corresponds to onset of the dissociation of Cooper pairs by sound where $\hbar\omega = 2\Delta(T_A)$, where $\hbar\omega$ is the energy of the sound quantum and $\hbar$ is Planck's constant. In the temperature range between T_A and the superfluid transition temperature, T_C , the attenuation of the liquid is extremely high owing to this mechanism. The point B corresponds to resonant absorption of sound at $\hbar\omega = 1.5\Delta(T)$ by the excited Cooper pairs with angular momentum $J = 2$ (ref. 8). Transverse sound is extinguished below this temperature. Early attempts to observe transverse sound in the normal phase of ^3He were inconclusive^{15–17}, and furthermore, it was originally expected that the transverse mode would be suppressed in the superfluid phase^{18,19}. More recent theoretical work⁸ clarified the role of the Cooper-pair excitations, showing that they increase the transverse sound speed which results in a more robust propagating transverse acoustic wave at low temperatures in superfluid $^3\text{He-B}$. The first experimental evidence for this can be found in the acoustic impedance measurements of ref. 20.

The proof that the impedance oscillations correspond to propagating transverse sound is given in Fig. 3. In Fig. 3a we show data sets obtained at a pressure of 4.42 bar in magnetic fields of 52 G, 101 G and 152 G. The principal feature is that the magnetic field modulates the zero-field oscillations shown in Fig. 2. Our detector is only sensitive to linearly polarized transverse sound having a specific direction. Application of a field of 52 G in the direction of wave propagation suppresses the oscillations near $T = 0.465 T_C$ that were present in zero field. This corresponds to a 90° rotation of the polarization of the first reflected transverse sound wave making the polarization orthogonal to the detection direction. Doubling the magnetic field restores the transverse sound oscillations at this temperature. The oscillations are suppressed once again by tripling the field to 152 G. We also note that near the points labelled 90° and 270° in Fig. 3, there are smaller-amplitude impedance oscillations

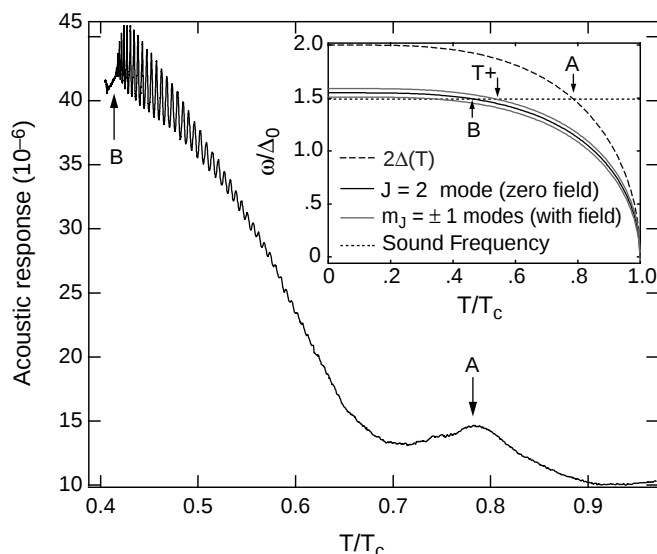


Figure 2 Temperature dependence of the acoustic cavity response. Measurements of the transverse acoustic impedance are shown at a pressure of 4.31 bar in zero magnetic field. The feature labelled A corresponds to acoustic pair-breaking, $\hbar\omega = 2\Delta(T_A)$. Impedance oscillations develop for temperatures $T < 0.6T_C$, and grow in amplitude at lower temperatures as the attenuation of transverse waves decreases. Near $T \approx T_B = 0.41T_C$ the propagating mode is extinguished by absorption of transverse sound via resonant excitation of Cooper pairs with total angular momentum $J = 2$ at $\hbar\omega = 1.5\Delta(T_B)$. The pair-breaking and resonance conditions of $J = 2$ Cooper-pair excitations for fixed frequency are shown in the inset; Δ_0 is the energy gap at zero temperature.

with shorter period than the primary oscillations. These come from interference of doubly reflected waves within the cavity. We demonstrate this fact with a simple, but powerful, simulation of the acoustic impedance oscillations (Fig. 3b; the results of the simulation are discussed later).

In zero field, superfluid $^3\text{He-B}$ is non-magnetic and non-birefringent. Linearly polarized transverse sound is the superposition of two circularly polarized waves having the same velocity and attenuation. Application of a magnetic field gives rise to acoustic circular birefringence through the Zeeman splitting of the excited states of the Cooper pairs that couple to the transverse sound modes: thus right- and left-circularly polarized waves propagate with different speeds, $C_\pm = C_0 \pm \delta C_\phi$. For magnetic fields well below 1 kG, the difference in propagation speeds is linear in the magnetic field, $\delta C_\phi \propto H$. This implies that a linearly polarized wave generated by the transducer undergoes Faraday rotation of its polarization as it propagates. On reflection from the opposite wall of the cavity, the linearly polarized wave with $\mathbf{q} \parallel \mathbf{H}$ reverses direction. The reflected wave propagates with the polarization rotating with the same handedness relative to the direction of the field; that is, the rotation of the polarization accumulates after reflection from a surface. The spatial period for rotation of the polarization by 360° is given by:

$$\Lambda = 4\pi \left(\frac{C_\phi}{\omega} \right) \left| \frac{C_\phi}{\delta C_\phi} \right| \quad (1)$$

The Faraday effect produces a sinusoidal modulation of the impedance oscillations as a function of magnetic field with a period that is inversely proportional to the field: that is, $\Lambda \propto 1/H$. The constant of proportionality in magneto-optics is called the Verdet constant, $V = 2\pi/\Lambda H$.

In Fig. 3b we show the result of our numerical calculation of the sound wave amplitude in the direction detected by the transducer. The oscillations shown in the figure come from interference between the source wave and multiply reflected waves. The calculation uses the attenuation and phase velocity measured in zero field.

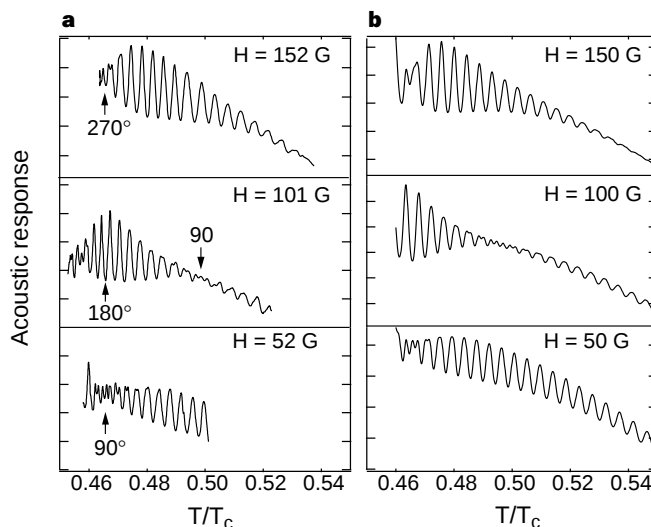


Figure 3 Magnetic field dependence of the acoustic cavity response. **a**, The sound amplitude measured at 4.42 bar. The angles are indicated for rotation of the polarization of transverse sound by 90° , 180° and 270° . **b**, Simulation of the acoustic impedance for the same path length cell and same temperature, pressure and magnetic fields. The wavelength and attenuation of transverse sound in zero field were used as the input data for the calculation. The Verdet constant for the simulation was determined by the position of the minimum in the impedance oscillations at $H = 52$ G shown in **a**.

The Verdet constant is obtained from the measurement at 52 G. The simulation reproduces all the observed features of the impedance as a function of temperature, including the maximum in the modulation at $T/T_c = 0.415$, $H = 101$ G, and the minimum at $T/T_c = 0.415$, $H = 152$ G, which confirms that the Faraday period is proportional to $1/H$. The simulation also produces the fine-structure oscillations in the impedance near the points labelled 90° and 270° . The fine structure is observed when the polarization rotates by an odd multiple of 90° upon a single round trip in the cell. Then waves that traverse the cell twice are 180° out of phase relative to the source wave, and consequently the period of the impedance oscillations is halved. The amplitude of the oscillations is substantially reduced because of attenuation over the longer path-length. This structure provides proof that impedance oscillations are modulated by the Faraday effect for propagating transverse waves. The impedance data from our experiments were analysed to obtain the spatial period for rotation of the polarization, and were found to be in agreement with the theoretical prediction⁸ for the Faraday rotation period. The theoretical results for the period can be expressed in the form

$$\Lambda = K \frac{\sqrt{T/T_c - 1}}{gH} \quad (2)$$

for fields $H \ll 1$ kG and temperatures above and near the extinction point B. The temperature T_+ corresponds to the extinction of transverse sound by resonant excitation of Cooper pairs with $J = 2$, $m_J = +1$, at a slightly higher temperature than the B extinction point in zero field as shown in Fig. 2 (for example, at $H = 100$ G, $T_+ - T_B \approx 1$ μ K). The magnitude of the Faraday rotation period depends on accurately known superfluid properties, contained in the parameter K ; it also depends on one parameter that is not well-established, the Landé g -factor, g , for the Zeeman splitting of the Cooper-pair excited state with $J = 2$.

Movshovich *et al.*²¹ analysed the splitting of the $J = 2$ multiplet in the absorption spectrum of longitudinal sound, finding a value of $g = 0.042$. In that experiment it was not possible to resolve the splitting except for fields above 2 kG. At these high fields, the nonlinear field dependence due to the Paschen–Back effect^{22,23} becomes comparable to the linear Zeeman splitting^{24,25}, which makes it difficult to determine the Landé g -factor accurately. We have analysed our measurements of the acoustic Faraday effect to determine the g -factor with high accuracy at low fields, which eliminates the complication of the high-field Paschen–Back effect. We find $g = 0.020 \pm 0.002$. We interpret our significantly smaller value of the Landé g -factor as meaning that there are important $L = 3$ (' f -wave') pairing correlations in the superfluid condensate, about 7% of the dominant p -wave interactions²⁶. $\square$

Received 8 February; accepted 10 June 1999.

- Landau, L. D. Oscillations in a Fermi liquid. *Sov. Phys. JETP* **5**, 101–108 (1957).
- Leggett, A. J. A theoretical interpretation of the new phases of liquid ^3He . *Rev. Mod. Phys.* **47**, 331–414 (1975).
- Vollhardt, D. & Wölfle, P. *The Superfluid Phases of Helium 3* (Taylor & Francis, New York, 1990).
- Anderson, P. W. & Brinkman, W. F. in *The Helium Liquids* (eds Armitage, J. G. M. & Farquhar, I. E.) 315–416 (Academic, New York, 1975).
- Wheatley, J. C. Experimental properties of superfluid ^3He . *Rev. Mod. Phys.* **47**, 415–470 (1975).
- Bardeen, J., Cooper, L. N. & Schrieffer, R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
- Maki, K. Propagation of zero sound in the Balian–Werthamer state. *J. Low Temp. Phys.* **16**, 465–477 (1974).
- Moore, G. F. & Sauls, J. A. Transverse waves in superfluid ^3He . *J. Low Temp. Phys.* **91**, 13–37 (1993).
- Avenel, O., Varoquaux, E. & Ebisawa, H. Field splitting of the new attenuation peak in ^3He . *Phys. Rev. Lett.* **45**, 1952–1955 (1980).
- Schopohl, N. & Tewordt, L. Landé factors of collective mode multiplets in ^3He -B and coupling strengths to sound waves. *J. Low Temp. Phys.* **45**, 67–90 (1981).
- Bennett, H. S. & Stern, E. A. Faraday and Kerr effects in ferromagnetics. *Phys. Rev.* **97**, A448–A461 (1965).
- Lee, Y. *et al.* High frequency acoustic measurements in liquid ^3He -B near the transition temperature. *J. Low Temp. Phys.* **103**, 265–272 (1996).
- Wölfle, P. & Einzel, D. Transport and relaxation properties of superfluid ^3He . II. *J. Low Temp. Phys.* **32**, 39–56 (1978).
- McKenzie, R. H. & Sauls, J. A. in *Helium Three* (eds Halperin, W. P. & Pitaevskii, L. P.) 255–311 (Elsevier Science, Amsterdam, 1990).

- Roach, P. R. & Ketterson, J. B. Observation of transverse zero sound in normal ^3He . *Phys. Rev. Lett.* **36**, 736–740 (1976).
- Flowers, E. G., Richardson, R. W. & Williamson, S. J. Transverse zero sound in normal ^3He . *Phys. Rev. Lett.* **37**, 309–311 (1976).
- Flowers, E. G. & Richardson, R. W. Transverse acoustic impedance of normal ^3He . *Phys. Rev.* **17**, 1238–1248 (1978).
- Combescot, M. & Combescot, R. Transverse zero sound propagation in superfluid ^3He . *Phys. Lett. A* **58**, 181–182 (1976).
- Maki, K. & Ebisawa, H. Transverse zero sound in superfluid ^3He . *J. Low Temp. Phys.* **26**, 627–636 (1977).
- Kalbfeld, S., Kucera, D. M. & Ketterson, J. B. Observation of an evolving standing-wave pattern involving a transverse disturbance in superfluid ^3He . *Phys. Rev. Lett.* **71**, 2264–2267 (1993).
- Movshovich, R., Varoquaux, E., Kim, N. & Lee, D. M. Splitting of the squashing collective mode of superfluid ^3He -B by a magnetic field. *Phys. Rev. Lett.* **61**, 1732–1735 (1988).
- Schopohl, N., Warnke, M. & Tewordt, L. Effect of gap distortion on the field splitting of collective modes in superfluid ^3He . *Phys. Rev. Lett.* **50**, 1066–1069 (1983).
- Shivaram, B. S., Meisel, M. W., Sarma, B. K., Halperin, W. P. & Ketterson, J. B. Nonlinear Zeeman shifts in the collective-mode spectrum of ^3He . *Phys. Rev. Lett.* **50**, 1070–1072 (1983).
- Halperin, W. P. & Varoquaux, E. in *Helium Three* (eds Halperin, W. P. & Pitaevskii, L. P.) 353–522 (Elsevier Science, Amsterdam, 1990).
- Movshovich, R., Varoquaux, E., Kim, N. & Lee, D. M. Fivefold splitting of the squashing mode of superfluid ^3He -B by a magnetic field. *Phys. Rev. B* **44**, 332–340 (1991).
- Sauls, J. A. & Serene, J. W. Interaction effects on the Zeeman splitting of collective modes in superfluid ^3He . *Phys. Rev. Lett.* **49**, 1183–1186 (1982).

Acknowledgements. We acknowledge contributions from J. Kycia and G. Moores, and support from the National Science Foundation and the NEDO Foundation of Japan.

Correspondence and requests for materials should be addressed to W.P.H. (e-mail: w-halperin@nwu.edu).

Similarities between the growth dynamics of university research and of competitive economic activities

Vasiliki Plerou*†, Luís A. Nunes Amaral*, Parameswaran Gopikrishnan*, Martin Meyer* & H. Eugene Stanley*

* Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

† Department of Physics, Boston College, Chestnut Hill, Massachusetts 02167, USA

Quantifying the dynamics of research activities is of considerable current interest, not least because of recent changes in research and development (R&D) funding^{1–9}. Here we quantify and analyse university research activities, and compare their growth dynamics with those of business firms^{10–14}. Our study involves the analysis of five distinct databases, the largest of which is a National Science Foundation database of the R&D expenditures in science and engineering for a 17-year period (1979–95) in 719 United States universities. We find that the distribution of growth rates displays a 'universal' form that does not depend on the size of the university or on the measure of size used; and the width of this distribution decays with size as a power law. These findings are quantitatively similar to those of business firms^{10–14}, and so are consistent with the hypothesis that the growth dynamics of complex organizations are governed by universal mechanisms. One possible explanation for these similarities is that the combination of peer review and government direction leads to an outcome similar to that induced by market forces (where the analogues of peer review and government direction are, respectively, consumer evaluation and product regulation).

In the study of physical systems, the scaling properties of fluctuations in the output of a system often yield information regarding the underlying processes responsible for the observed macroscopic behaviour^{15,16}. Here we analyse the fluctuations in the growth rates of university research activities, using five different measures of research activity. The first measure of the size of a university's research activities that we consider is R&D expenditure.

The rationale for using this as a measure of research activity is that research is an expensive activity that the university finances with external support.

We first analyse a database containing the annual R&D expenditure for science and engineering of 719 US universities¹⁷ for the 17-year period 1979–95 (~12,000 data points). The expenditures are broken down by school and department. The annual growth rate of R&D expenditures is, by definition, $g(t) \equiv \log[S(t+1)/S(t)]$, where $S(t)$ and $S(t+1)$ are the R&D expenditures of a given university in the years t and $t+1$, respectively. We expect that the statistical properties of the growth rate g depend on S , as it is natural that the fluctuations in g will decrease with S . Therefore, we partition the universities into groups according to the size of their R&D expenditure (Fig. 1a). Figure 1b suggests that the conditional probability density, $p(g|S)$, has the same functional form, with different widths, for all S .

We next calculate the width $\sigma(S)$ of the distribution of growth rates as a function of S . Figure 1c shows that $\sigma(S)$ scales as a power law

$$\sigma(S) \propto S^{-\beta} \quad (1)$$

with $\beta = 0.25 \pm 0.05$. In Fig. 1d, we collapse the scaled conditional probability distributions onto a single curve.

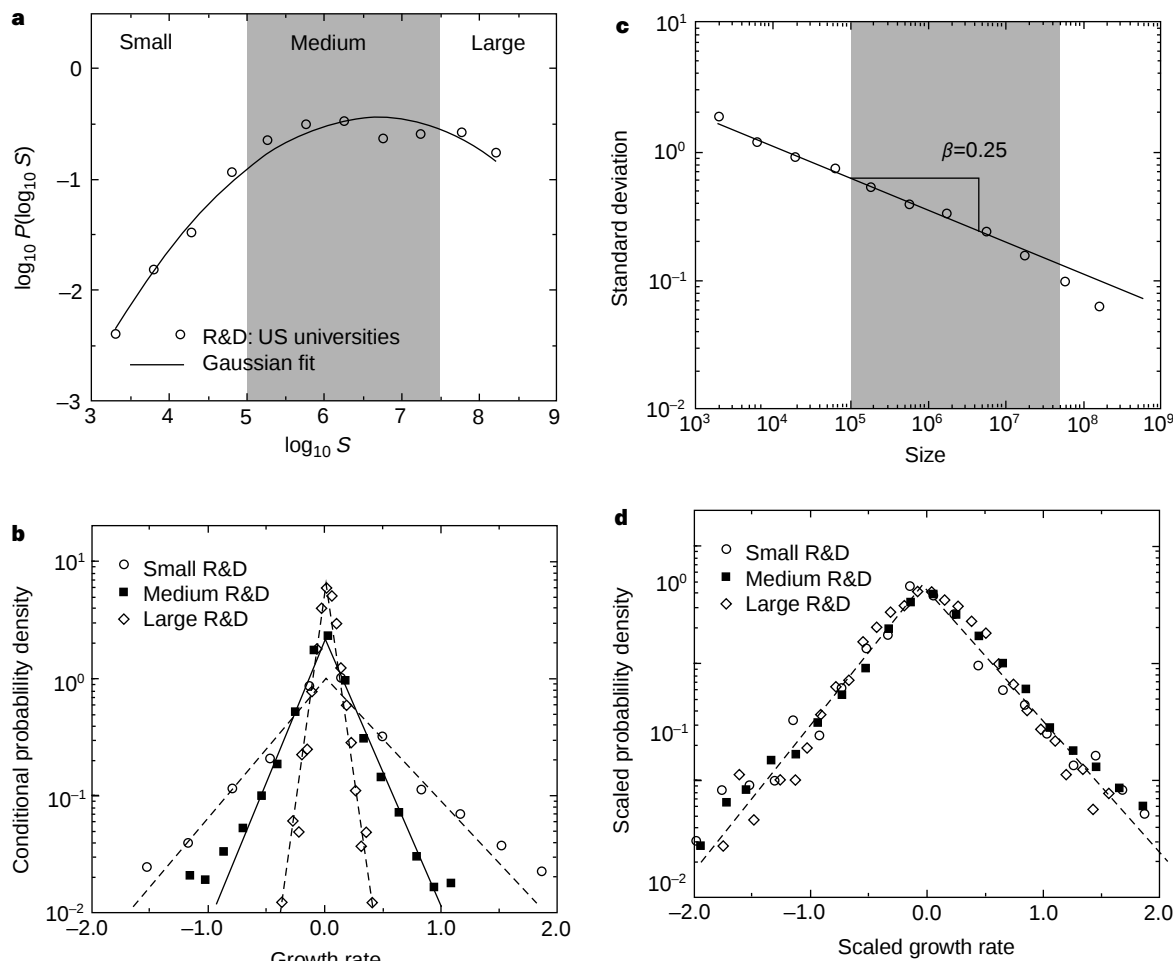


Figure 1 Growth dynamics of research activities at universities. **a**, Histogram of the logarithm of the annual R&D expenditure of 719 US universities for the 17-year period 1979–95, expressed in 1992 US dollars. Here S denotes the R&D expenditures detrended by inflation so that values for different years are comparable. The bins were chosen equally spaced on a logarithmic scale with bin size 0.5. The line is a gaussian fit to the data, which is a prediction of Gibrat's theory^{10,12}. **b**, Conditional probability density function $p(g|S)$ of the annual growth

To test if these results for the dynamics of R&D expenditures are valid for other measures of research activity, we next analyse another measure of a university's research activities, the number of papers published each year^{18–20}. We analyse data for the 17-year period 1981–97 from ref. 21, which records the number of papers published by the top 112 US universities (~1,900 data points). We find that the analogue of Fig. 1 holds. We note that the same exponent value, $\beta = 1/4$, is found (Fig. 2a) and that the same functional form of $p(g|S)$ is displayed (Fig. 2b).

Next, we consider as a measure of size the number of patents issued to a university²². We retrieve from ref. 23 the number of patents issued to each of 106 universities each year of the 22-year period 1976–97 (~2,300 data points). We confirm that the analogue of Fig. 1 holds, with the same exponent value, $\beta = 1/4$ (Fig. 2a), and the same functional form of $p(g|S)$, Fig. 2b.

To test if our findings hold for different academic systems, we analyse two databases on research funding of English²⁴ and Canadian²⁵ universities. The same quantitative behaviour is found for the distribution of growth rates and for the scaling of σ , with the same exponent value (Fig. 2a) and the same functional form of $p(g|S)$, Fig. 2b. Thus, the analysis of all five databases confirms that the same quantitative results hold across different measures of research activity and academic systems.

rates g . For this plot the entire database is divided into three groups (depicted in **a** by different shades). **c**, Standard deviation $\sigma(S)$ of the distribution of annual growth rates as a function of S . The straight line is a power-law fit to the data, and its slope gives the exponent $\beta = 0.25 \pm 0.05$. **d**, Scaled probability density function $p(g|S)/\sigma^{-1}(S)$ plotted against the scaled annual growth rate $(g - \bar{g})/\sigma(S)$ for the three groups defined in **b**. Note that the scaled data collapse onto a single curve.

We next address the question of how to interpret our empirical results. We start with the observation that research is an expensive activity, and that the university must 'offer' its research to external sources such as governmental agencies and business firms. Thus, an increase in R&D expenditures at university A and a decrease at university B implies that the funders of research increasingly choose their research from university A as opposed to university B¹. This qualitative picture parallels the competition among different business firms, so it is natural to enquire if there is quantitative support for this analogy between university research and business activities. To quantitatively test this analogy, we note that the results of Fig. 1 are remarkably similar to the results found for firms^{13,14} and countries^{26–28}. We plot in Fig. 2c the scaled conditional probabilities $p(g|S)$ for countries, firms and universities, and find that the distributions for the different organizations fall onto a single curve.

There is, however, one difference: for firms and countries, we find $\beta \approx 1/6$, while for universities, $\beta \approx 1/4$. We can understand this difference using a model for organization growth²⁹. In the model, each organization—university, firm or country—is made up of units. The model assumes these units grow through an independent,

gaussian-distributed, random multiplicative process with variance W^2 . Units are absorbed when they become smaller than a 'minimum size', which is a function of the activity they perform. Units can also give rise to new units if they grow by more than the minimum size for a new unit to form. The model predicts $\beta = W/[2(W + D)]$, where D is the width of the distribution of minimum sizes for the units²⁹. For firms, the range of typical sizes is very broad—from small software and accounting firms to large oil and car firms—suggesting a large value of D . On the other hand, for universities, the range of typical sizes is much narrower, suggesting a small value of D and implying a larger value of β than for business firms. This is indeed what we observe empirically.

Business firms are comprised of divisions and universities are made up of schools or colleges, so it is natural to consider the internal structure of these complex organizations³⁰. We next quantify how the internal structure of a university depends on its size by calculating the conditional probability density $\rho(\xi|S)$ to find a school of size ξ in a university of size S (Fig. 3a). The model predicts that $\rho(\xi|S)$ obeys the scaling form²⁹

$$\rho(\xi|S) \propto S^{-\alpha} f(\xi/S^\alpha) \quad (2)$$

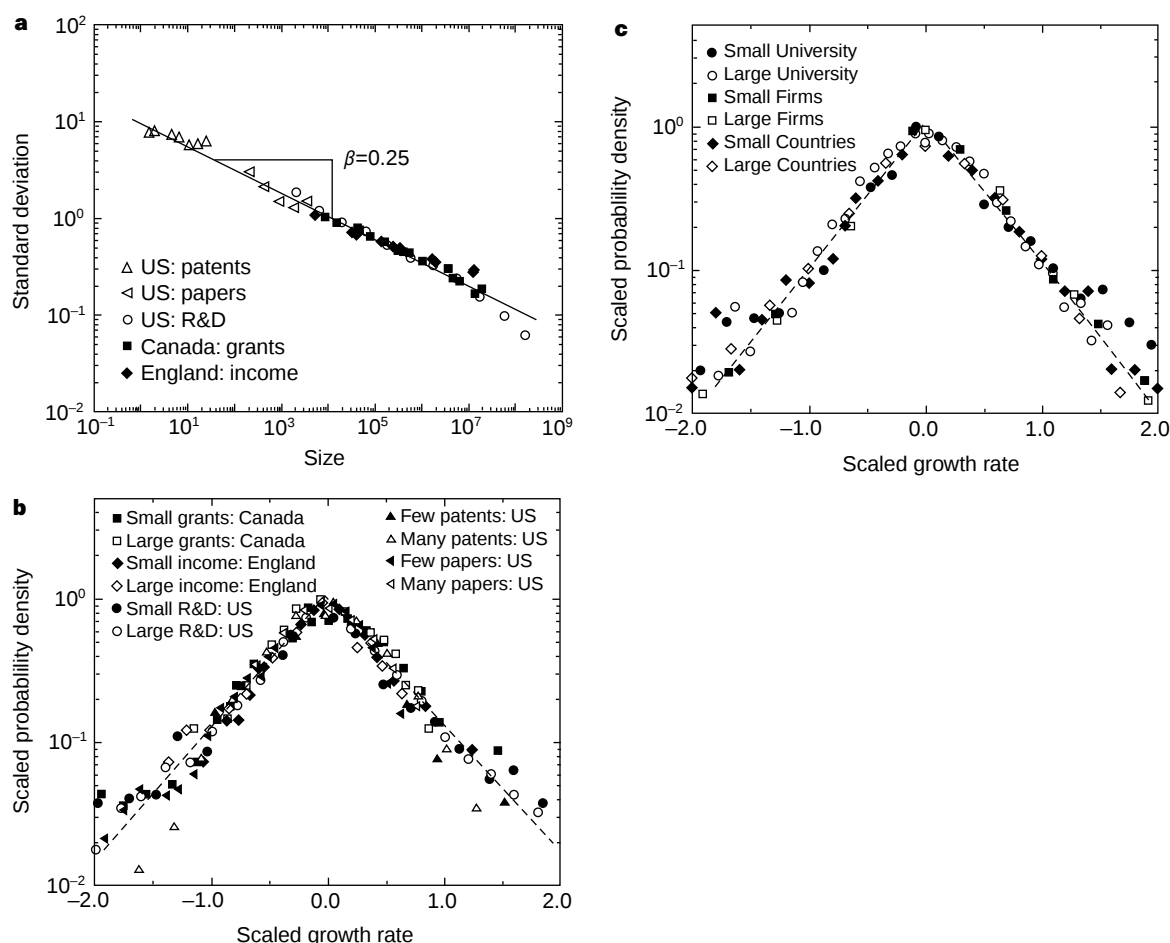


Figure 2 Robustness of empirical findings for the distribution of growth rates. **a**, Standard deviation $\sigma(S)$ of the distribution of annual growth rates for different measures of research activities and different academic systems from the data in the five distinct databases analysed: (1) the number of papers published each year at 112 US universities, (2) the number of patents issued each year to 106 universities, (3) the R&D expenditures in US dollars of 719 US universities, (4) the total amount in Canadian dollars of the grants to 60 Canadian universities, and (5) the external incomes in British pounds of 90 English universities. It is apparent that for all measures and all academic systems analysed, we find a power-law dependence—with the same exponent, $\beta \approx 1/4$. The values of σ for the different

measures were shifted vertically for better comparison of the estimates of the exponents. **b**, The distribution of annual growth rates, scaled as in Fig. 1d, for the five databases. We show the distribution of growth rates for two different groups, obtained in a way similar to that described in Fig. 1b, for each of the five measures. The data appear to collapse onto a single curve, suggesting that the different measures have similar statistical properties. **c**, The distribution of scaled annual growth rates for different organizations: R&D expenditures of US universities, sales of firms, and GDP of countries. The data collapse onto a single curve, suggesting that the scaled distributions have the same functional form.

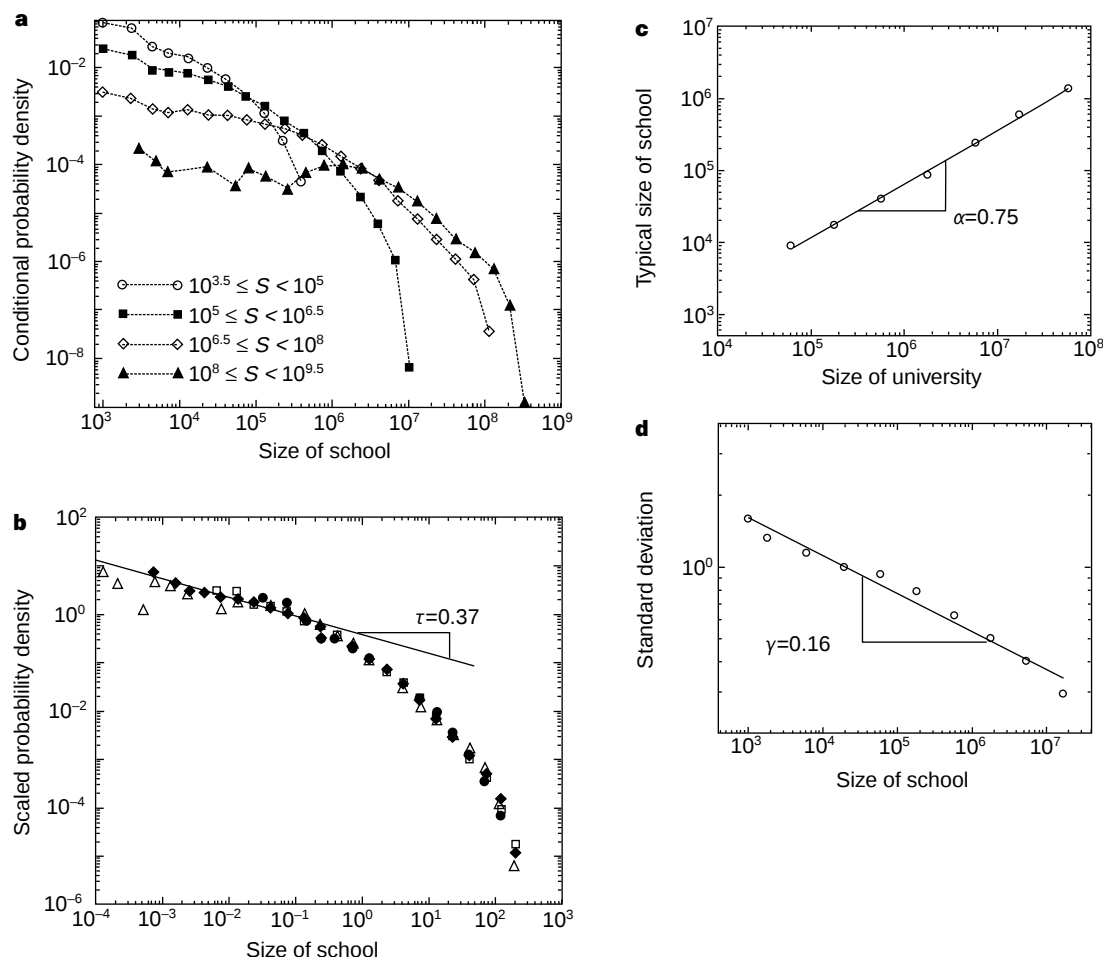


Figure 3 Statistical analysis of the units forming the internal structure of a university, the schools. **a**, Conditional probability function $\rho(\xi|S)$ of finding a school of size ξ in a university of size S . To improve the statistics, we partition the universities by size into four groups. **b**, To illustrate the scaling relation (equation (2)), we plot the scaled probability density $\rho(\xi|S)/S^{-\alpha}$ versus the scaled size of the school ξ/S^α . In agreement with equation (2), we find that the scaled

data fall onto a single curve. **c**, Scaling of the typical size of a school in a university of a given size for different university sizes. The data obey a power law with exponent $\alpha = 0.75 \pm 0.05$. **d**, Standard deviation W of the distribution of growth rates of schools versus school size ξ . The data obey a power law with exponent $\gamma = 0.16 \pm 0.05$. Using equation (4) and this value of γ , we obtain an independent estimate $\beta = 0.25 \pm 0.05$.

where $f(u) \approx u^{-\tau}$ for $u \ll 1$, and $f(u)$ decays as a stretched exponential for $u \gg 1$. We find $\tau = 0.37 \pm 0.10$ (Fig. 3b), and $\alpha = 0.75 \pm 0.05$ (Fig. 3c). We test the scaling hypothesis (equation (2)) by plotting the scaled variables $\rho(\xi|S)/S^{-\alpha}$ versus ξ/S^α . Figure 3b shows that all curves collapse onto a single curve, which is the scaling function $f(u)$.

Equation (2) implies that the typical number of schools with research activities in a university of size S scales as $S^{1-\alpha}$, while the typical size of these schools scales as S^α . Hence, we can calculate how σ depends on S :

$$\sigma(S) \propto (S^{1-\alpha})^{-1/2} W(\xi) \quad (3)$$

To determine σ , we first find the dependence of W on ξ . Figure 3d shows that $W \propto \xi^{-\gamma}$ with $\gamma = 0.16 \pm 0.05$. Substituting into equation (3) and remembering that the typical size of the schools is S^α , we obtain $\sigma(S) \propto (S^{1-\alpha})^{-1/2} (S^\alpha)^{-\gamma}$, which leads to the testable exponent relation:

$$\beta = \frac{1-\alpha}{2} + \alpha\gamma \quad (4)$$

For $\alpha \approx 3/4$ and $\gamma \approx 1/6$, equation (4) predicts $\beta \approx 1/4$, in agreement with our empirical estimate of β from the five distinct databases analysed (Fig. 2a).

Our results are consistent with the possibility that the statistical

properties of university research activities are surprisingly similar for different measures of research activity and for distinct academic systems. Moreover, our findings for university research resemble those independently found for business firms^{10–14} and countries^{26–28}. One possible explanation is that peer review, together with government direction, may lead to an outcome similar to that induced by market forces, where the analogue of peer-review quality control may be consumer evaluation, and the analogue of government direction may be product regulation.

Practical implications of our finding of similar growth dynamics for academic research and business are not obvious, but one possibility is that there seems to be no need to make academic research at universities still more like business—it already is. Some may claim that the business sector could be regarded as a yardstick for organizing academic research. If so, the research departments already behave like business units and hence are sufficiently ‘effective’. On the other hand, others may maintain that the ‘economization’ of academic research has been pushed too far, and that the research system will become ‘ineffective’ if this continues. □

Received 5 April; accepted 14 June 1999.

1. Moed, H. F., Luwel, M., Houben, J. A., Van Den Berghe, H. & Spruyt, E. Funding and research performance. *Nature* **392**, 119 (1997).
2. Mifflin, B. Institutes top UK science league table. *Nature* **390**, 12 (1997).
3. Buffetaut, E. No independence for French researchers. *Nature* **392**, 542 (1998).

4. Anderson, C. Job security—Tenure, under fire once again, still holds strong. *Science* **265**, 1923 (1994).
5. Holden, C. Tenure turmoil sparks reforms. *Science* **276**, 24–26 (1997).
6. Halpern, J. J. & Velleman, P. F. Tenure tracking. *Science* **276**, 1321 (1997).
7. Geissman, J. W. *et al.* Tenure tracking. *Science* **276**, 1320–1321 (1997).
8. Gazzaniga, M. S. How to change the university. *Science* **282**, 237 (1998).
9. Cooper, W. E. Restructuring the university. *Science* **282**, 1047 (1998).
10. Gibrat, R. *Les Inégalités Économiques* (Sirey, Paris, 1931).
11. Ijiri, Y. & Simon, H. A. *Skew Distributions and the Sizes of Business Firms* (North Holland, Amsterdam, 1977).
12. Sutton, J. Gibrat's legacy. *J. Econ. Lit.* **35**, 40–59 (1997).
13. Stanley, M. H. R. *et al.* Scaling behaviour in the growth of companies. *Nature* **379**, 804–806 (1996).
14. Takayasu, H. & Okuyama, K. Country dependence on company size distributions and a numerical model based on competition and cooperation. *Fractals* **6**, 67–79 (1998).
15. Vicsek, T. *Fractal Growth Phenomena* 2nd edn (World Scientific, Singapore, 1992).
16. Bunde, A. & Havlin, S. *Fractals and Disordered Systems* (Springer, Berlin, 1991).
17. National Science Foundation, Division of Science Resources Studies *Academic Research and Development Expenditures* (NSF, Arlington, Virginia, 1998).
18. Braun, T. & Schubert, A. Indicators of research output in the sciences for 5 central European countries. *Scientometrics* **36**, 145–165 (1996).
19. Lewison, G. New bibliometric techniques for the evaluation of medical schools. *Scientometrics* **41**, 5–16 (1998).
20. Schwarz, A. W., Schwarz, S. & Tijssen, R. J. W. Research and research impact of a technical university—A bibliometric study. *Scientometrics* **41**, 371–388 (1998).
21. *United States University Science Indicators on Diskette, 1981–1997* (Inst. for Scientific Information, Philadelphia, 1998).
22. Narin, F. Patents as indicators for the evaluation of industrial research output. *Scientometrics* **34**, 489–496 (1995).
23. United States Patent and Trademarks Office Databases 1976–1997 (<http://www.uspto.gov>).
24. Higher Education Funding Council for England *The 1996 Research Assessment Exercise* (HEFCE, Bristol, 1996).
25. Natural Sciences and Engineering Research Council of Canada *NSERC Grant Database for 1991–1998* (NSERC, Ottawa, 1999).
26. Summers, R. & Heston, A. The Penn World Tables (Mark 5): An expanded set of international comparisons, 1950–1988. *Q. J. Econ.* **106**, 327–368 (1991).
27. Durlauf, S. N. On the convergence and divergence of growth rates. *Econ. J.* **106**, 1016–1018 (1996).
28. Lee, Y., Amaral, L. A. N., Canning, D., Meyer, M. & Stanley, H. E. Universal features in the growth dynamics of complex organizations. *Phys. Rev. Lett.* **81**, 3275–3278 (1998).
29. Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Salinger, M. A. & Stanley, H. E. Power law scaling for a system of interacting units with complex internal structure. *Phys. Rev. Lett.* **80**, 1385–1388 (1998).
30. Jovanovic, B. The diversification of production. *Brookings Pap. Econ. Activity: Microeconomics* (1) 197–247 (1993).

Acknowledgements. We thank M. Barthélemy, S. V. Buldyrev, D. Canning, X. Gabaix, S. Havlin, P. Ch. Ivanov, H. Kallabis, Y. Lee and B. Roehner for discussions. We also thank N. Bayers, E. Garfield, and especially R. E. Hudson for help with obtaining the ISI database, and H. F. Moed for comments on the manuscript, and in particular for suggesting the arguments about effectiveness used in the last paragraph. This work was supported by the NSF L.A.N.A. thanks FCT/Portugal for support.

Correspondence and requests for materials should be addressed to H.E.S. (e-mail: HES@bu.edu).

Abrupt changes in North American climate during early Holocene times

F. S. Hu*, D. Slawinski†, H. E. Wright Jr†, E. Ito†, R. G. Johnson†, K. R. Kelts†, R. F. McEwan† & A. Boedigheimer†

* Departments of Plant Biology and Geology, University of Illinois, Urbana, Illinois 61801, USA

† Limnological Research Center and Department of Geology & Geophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA

Recent studies of the Greenland ice cores have offered many insights into Holocene climatic dynamics at decadal to century timescales^{1–3}. Despite the abundance of continental records of Holocene climate, few have sufficient chronological control and sampling resolution to compare with the Greenland findings⁴. But annually laminated sediments (varves) from lakes can provide high-resolution continental palaeoclimate data with secure chronologies. Here we present analyses of varved sediments from Deep Lake in Minnesota, USA. Trends in the stable oxygen-isotope composition of the sedimentary carbonate indicate a pronounced climate cooling from 8.9 to 8.3 kyr before present, probably characterized by increased outbreaks of polar air, decreased precipitation temperatures, and a higher fraction of the annual precipitation falling as snow. The abrupt onset of this climate reversal, over several decades, was probably caused by a

reorganization of atmospheric circulation and cooling of the Arctic airmass in summer that resulted from the final collapse of the Laurentide ice near Hudson Bay and the discharge of icebergs from the Quebec and Keewatin centres into the Tyrell Sea. The timing and duration of this climate reversal suggest that it is distinct from the prominent widespread cold snap that occurred 8,200 years ago in Greenland and other regions^{1,5,6}. No shifts in the oxygen-isotope composition of sediment carbonate occurred at 8.2 kyr before present at Deep Lake, but varve thickness increased dramatically, probably as a result of increased deposition of aeolian dust. Taken together, our data suggest that two separate regional-scale climate reversals occurred between 9,000 and 8,000 years ago, and that they were driven by different mechanisms.

We retrieved three temporally overlapping sediment cores from Deep Lake (47° 41' N, 95° 23' W; Fig. 1), a topographically closed basin located ~45 km east of the prairie–forest border in north-western Minnesota. The early Holocene sediments of these cores are varve couplets consisting of calcite precipitated through photosynthesis in summer paired with darker, clastic and organic debris deposited in other seasons. Annual laminae are easily identifiable under a low-power microscope, providing a high-quality chronometer. To minimize the accumulative error of varve counting from the surface, we anchored our varve counts on a reliable AMS (accelerator mass spectrometry) ¹⁴C date of 8,090 ± 85 years (calibrated to 8,986 kyr calendar age⁷) on a large wood sample⁸. A second AMS ¹⁴C date of 8,740 ± 60 years, also on a piece of wood, is supportive of the annually laminated nature of early Holocene sediments from Deep Lake, but this date is less useful because it is on a ¹⁴C plateau and yields ambiguous calibrated ages⁷. Varve thickness was measured, and subsamples from the cores were analysed for stable isotopes at multi-decadal resolution to reconstruct early Holocene climate. Our results provide an opportunity to test whether decadal to century scale climate changes observed in Greenland ice also occurred near the centre of the North American continent.

Bulk-carbonate $\delta^{18}\text{O}$ of the Deep Lake core shows distinct stratigraphic changes between 10.0 and 8.0 calendar kyr before present (cal. kyr BP; Fig. 2a). $\delta^{18}\text{O}$ increases by ~2‰ about 9.5 cal. kyr BP, reflecting the effects of regional climate warming related to increased summer insolation and the retreat of the Laurentide ice sheet⁴, as well as the associated increase in evaporation from Deep Lake itself and the reduced influence on climate of glacial Lake Agassiz as it retreated into Canada^{8,9}. Around 8.9 cal. kyr BP, $\delta^{18}\text{O}$

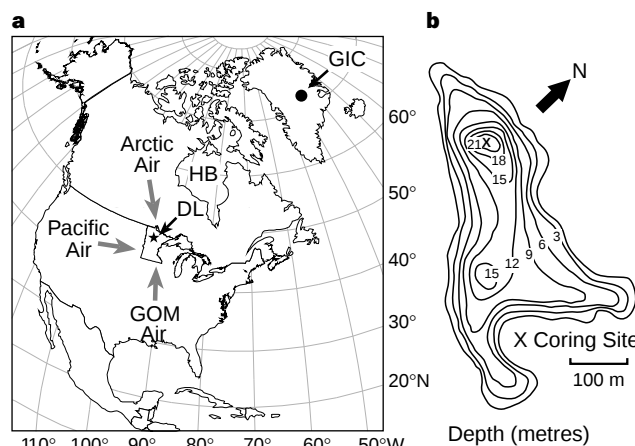


Figure 1 Air masses, site locations and lake bathymetry. **a**, General directions of the three main air masses controlling the climate of Minnesota (GOM, Gulf of Mexico); also shown are the locations of Deep Lake (DL), Greenland ice cores (GIC), and Hudson Bay (HB). **b**, Bathymetry of Deep Lake.

declines abruptly by $\sim 2\text{‰}$ within <50 yr. It remains low at about -8‰ until 8.4 cal. kyr BP, when it gradually increases to $\sim -6.5\text{‰}$ by 8.2 cal. kyr BP and stays constant at least until 7.8 cal. kyr BP. After 9.0 cal. kyr BP, $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ significantly co-vary ($R^2 = 0.820$, $p < 0.001$, $n = 42$) with similar stratigraphic patterns.

The abrupt decrease in $\delta^{18}\text{O}$ at 8.9 cal. kyr BP post-dates the forest–prairie transition at Deep Lake at ~ 9.5 cal. kyr BP⁸, but it coincides with an overall increase in aeolian deposition inferred from increased varve thickness (Fig. 3), as discussed below. This coincidence suggests the possibility that input of detrital carbonate caused the observed isotopic changes. However, no consistent differences in either $\delta^{18}\text{O}$ or $\delta^{13}\text{C}$ exist between samples of thick-varve and thin-varve bundles within the period of interest (Fig. 2b). In addition, X-ray diffraction analysis of selected samples shows no corresponding changes in carbonate composition, and microscopic examination of sediments indicates the overall dominance of low-Mg authigenic calcite throughout the profile. Thus the observed isotopic changes are not caused by changes in carbonate composition due to detrital input.

Because lake-sediment $\delta^{18}\text{O}$ records from continental regions reflect a number of climatic and hydrologic factors^{10,11}, it is difficult to pinpoint the climatic causes of the $\delta^{18}\text{O}$ fluctuations at Deep Lake. However, this lake is today located in a climatically sensitive area near the junction of the Gulf-of-Mexico, Pacific and Arctic airmasses¹². The $\delta^{18}\text{O}$ reversal at 8.9 cal. kyr BP probably resulted from the interplay of these airmasses. A change of moisture source is unlikely to have been an important factor in determining $\delta^{18}\text{O}$ values, because both the Pacific and Arctic airmasses carry little moisture and thus the dominant moisture source was probably the Gulf of Mexico throughout the Holocene. Nevertheless, as at present in the midwestern United States, the diminished influence of the warm Gulf-of-Mexico airmass in favour of the Arctic airmass would probably result in more frequent polar outbreaks, lower precipitation temperatures, and a greater proportion of precipitation as snow^{12,13}. These climate changes should enhance ^{18}O depletion of precipitation^{10,11,13}, leading to the 2‰ decrease in $\delta^{18}\text{O}$ at Deep Lake.

The abrupt decline in $\delta^{18}\text{O}$ at ~ 8.9 cal. kyr BP coincides with the drainage at glacial lakes Agassiz and Ojibway to the Labrador Sea during the sudden collapse of the Hudson Bay ice dome^{14–17}, providing that the ^{14}C dates reported are calibrated⁷. The Hudson Bay ice dome surged repeatedly into glacial Lake Ojibway—the so-called Cochrane surges¹⁵—shortly before its collapse. As a result, the Hudson Bay ice dome disintegrated rapidly, and at 8.0 ^{14}C kyr BP (8.8 cal. kyr BP), lakes Agassiz and Ojibway drained northward catastrophically through the Hudson Strait^{14,16}. The drainage resulted in a flux of $\sim 1.2 \times 10^3 \text{ km}^3$ ($\sim 3.75 \text{ Sv}$ if the drainage occurred within one year) of fresh water into the northwestern Atlantic Ocean and left distinct traceable sediment units in the southeastern Hudson Bay area as well as throughout the Hudson Strait. This freshwater influx would be sufficient to slow deep-water formation¹⁸, thereby causing a significant reorganization in atmospheric circulation patterns¹⁹. One conceivable result of this reorganization would be that the polar front migrated southwards, and that the zonal flow was strengthened in the midwestern United States. These changes would have weakened the influence of the Gulf-of-Mexico airmass in the Midwest. In addition, after the collapse of the Hudson Bay ice dome, the Arctic airmass would have been cooled in summer as it passed over the Tyrell Sea, which was continually supplied by icebergs from the still active Quebec and Keewatin ice centres. The temperature gradient would have been steep from here to the mid-continent south and west of Minnesota, where temperatures were high because of greater summer insolation at that time. The jet stream and associated storm tracks must have been focused on this steep temperature/pressure gradient²⁰, and the cooled and strengthened Arctic air would prolong winter conditions. As a result, moist Gulf air entrained by storms passing over Deep Lake would have been

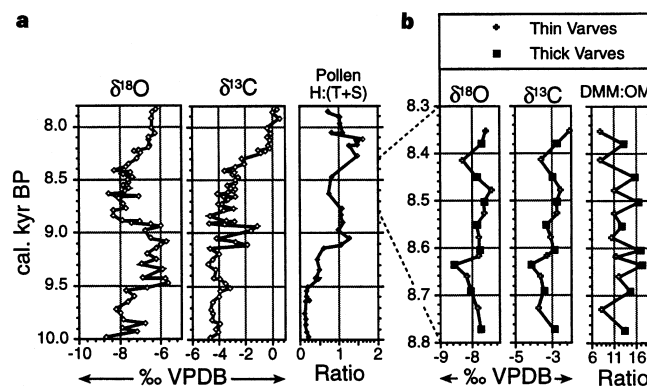


Figure 2 Proxy climate data from Deep Lake. **a**, Oxygen and carbon isotopes of bulk carbonates, and pollen ratios of total herbs to trees plus shrubs, H:(T+S), Deep Lake. **b**, Oxygen and carbon isotopes, and ratios of detrital mineral matter content (DMM) to organic matter content (OM) for thick-varve versus thin-varve bundles. For the analyses of oxygen and carbon isotopes, bulk-carbon samples were reacted at 25 °C for 1 h with 104% H_3PO_4 made from ultra-pure P_2O_5 and triply distilled H_2O (ref. 28); this reaction should not result in the substantial dissolution of detrital dolomite, which could complicate climatic interpretation of the data. The isotopic composition of the released CO_2 gas was analysed with a Finnigan MAT delta E triple-collector mass spectrometer. (The VPDB standard was used for the $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ values.) Pollen analysis followed standard procedures, and at least 300 pollen grains were counted at each level by K. Lease. Organic and carbonate contents were determined by loss on ignition at 500 and 1,000 °C; DDM = (100 – OM – carbonate content).

lifted to higher altitudes and cooled to a greater extent, releasing precipitation with lower $\delta^{18}\text{O}$ values and contributing to the 2‰ decline in the Deep Lake profile at 8.9 cal. kyr BP. The icebergs that calved into the Tyrell Sea from the residual Keewatin and Quebec ice domes¹⁶ probably sustained climate cooling until 8.3 cal. kyr BP.

If the Hudson Bay deglaciation caused large changes in atmospheric circulation that resulted in the $\delta^{18}\text{O}$ reversal at Deep Lake, similar climate changes were probably widespread. A recent synthesis of stratigraphic data¹⁹ indicates altered atmospheric circulation patterns at this time, as suggested by a significant decrease in marine productivity in the northwestern North Atlantic and by declined values of pollen influx and cellulose $\delta^{18}\text{O}$ in lake sediments from Quebec and adjacent New England. In the Greenland ice records, a number of proxies—including snow accumulation, $\delta^{18}\text{O}$ -derived temperature reconstruction, and Ca and CH_4 contents—also show¹ an anomaly lasting from 9.0 to 8.0 cal. kyr BP, which may also be related to these inferred changes in atmospheric circulation resulting from the draining of lakes Agassiz and Ojibway. However, no evidence has been found that salinity and sea surface temperatures changed at this time in the northwestern North Atlantic; this is either because the freshwater flux was insufficient to cause such changes or because existing data lack adequate chronological control and sampling detail.

Abundant Holocene records of continental climate changes are available from the midwestern United States²¹ and eastern North America²², some of which show compelling evidence for marked climate reversals during the early Holocene¹. Unfortunately, these previous records lack the dating accuracy and sampling detail necessary to test the geographical extent of the possible climate effects induced by the drainage of these large proglacial lakes. Stable-isotope analysis has been conducted on the varved sediments of Elk Lake (~ 60 km southeast of Deep Lake), Minnesota²³. However, the coarse resolution of the early Holocene $\delta^{18}\text{O}$ data from this site is unsuitable for testing the $\delta^{18}\text{O}$ reversal at ~ 8.9 –8.3 cal. kyr BP at Deep Lake. Nevertheless, a recent study of diatoms in a sediment

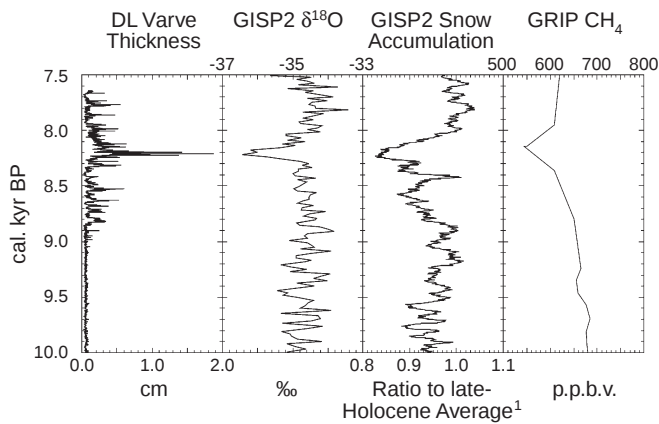


Figure 3 Palaeoclimate records from Deep Lake and Greenland. Varve-thickness data from Deep Lake are compared with records of $\delta^{18}\text{O}$ (refs 29, 30), snow accumulation ratio¹, and CH_4 concentration² from the Greenland ice. DL, Deep Lake; GISP2, US Greenland Ice-Sheet Project 2; GRIP, European Greenland Ice-Core Project. Note that the ^{14}C date on which the Deep Lake varve chronology is anchored has a standard error of ± 85 years and that the climate reversal that peaked at 8.2 cal. kyr BP lasted from 8.0 to 8.4 cal. kyr BP in the Greenland ice-core record¹. Varve thickness was measured with an Apple Macintosh PowerPC computer equipped with an Audio/Visual package and a Panasonic digital video camera to capture varve images. These images were then transformed with Photoshop version 2.5 (Adobe, San Jose, California, 1993) and digitized with Image (US Natl Institutes of Health, Bethesda, Maryland; available at <http://rsb.info.nih.gov/nih-image>).

core from Moon Lake (~ 100 km west of Deep Lake) in eastern North Dakota provides an early Holocene salinity record at multi-decadal resolution²⁴, the chronology of which is well constrained with multiple AMS ^{14}C dates of terrestrial plant materials. This record also shows a pronounced reversal at ~ 8.9 – 8.3 cal. kyr BP from the increasing salinity trend of the early Holocene. This salinity reversal was probably induced by the same climatic mechanism (particularly through decreased evaporation as a result of decreased summer temperature) responsible for the concurrent ^{18}O depletion at Deep Lake. In addition, our new initial results of high-temporal-resolution sediment analyses from Twin Lakes in northwestern Minnesota show an abrupt, pronounced decrease in $\delta^{18}\text{O}$ starting at ~ 9.0 cal. kyr BP, thus duplicating the Deep Lake record.

Coincident with the abrupt $\delta^{18}\text{O}$ decrease at 8.9 cal. kyr BP is an overall increase in varve thickness (Figs 2 and 3). Varve thickness in Minnesota lakes near the forest–prairie border is thought to be a sensitive recorder of aeolian activity²⁵. Increased varve thickness at ~ 9.0 cal. kyr BP probably reflects increased aeolian dust input to Deep Lake from the northern Great Plains to the west as a result of strengthened zonal flow. This interpretation is supported by the fact that thicker varves generally have higher ratios of detrital mineral to organic matter (Fig. 2) and is consistent with the climatic interpretation of the $\delta^{18}\text{O}$ record discussed above. This climatic effect was probably strongest in the mid-continent, which was closer to the Tyrell Sea than the North Atlantic region.

The onset of the $\delta^{18}\text{O}$ reversal 8.9 cal. kyr BP at Deep Lake significantly pre-dates the prominent climate event that occurred between 8.4 and 8.0 cal. kyr BP and peaked at 8.2 cal. kyr BP. Evidence for this event has been found in various proxy records from the Greenland ice and other parts of the globe^{1,5,6}. Nevertheless, a prominent peak in varve thickness occurs at 8.2 cal. kyr BP at Deep Lake (Fig. 3). Within the possible chronological errors of the Deep Lake and Greenland records, this varve-thickness peak correlates with changes in various climate proxies (for example, $\delta^{18}\text{O}$, snow accumulation, and CH_4 concentration) in the Greenland ice,

providing strong evidence that the 8.2-cal. kyr event affected regions far from Greenland. If this event resulted from a temporary perturbation in thermohaline circulation in the North Atlantic circulation¹, it may have had a widespread effect on climate, causing enhanced drought intensity and aeolian activity in the North American mid-continent, as suggested by increased varve thickness and high pollen ratios of herbs to trees plus shrubs at Deep Lake (Fig. 2). Such an interpretation would seem to be supported by increased varve thickness at this time in the sediments of Elk Lake, Minnesota²⁵, although the early Holocene chronology of this record may be compromised by a substantial accumulative error²⁶ related to varve counting from the sediment–water interface. Values of $\delta^{18}\text{O}$ do not show oscillations corresponding to the varve-thickness peak at 8.2 cal. kyr BP at Deep Lake, implying that the controls are different from those for the reversal 8.9–8.3 cal. kyr BP. One possible explanation for the lack of an isotopic response at 8.2 cal. kyr BP is that $\delta^{18}\text{O}$ depletion due to climate cooling and associated reduced evaporation^{10,11,13} was compensated by evaporative ^{18}O enrichment resulting from dry-climate conditions. As inferred from the much thicker varves and higher pollen ratios of herbs to trees plus shrubs at 8.2 cal. kyr BP than at 8.9–8.3 cal. kyr BP, drought intensity was probably much greater at 8.2 cal. kyr after the cooling effect of icebergs in the Tyrell Sea terminated, which may have contributed to the difference of $\delta^{18}\text{O}$ signals between the two climate events.

The mechanism of the 8.2-cal. kyr BP event is a subject of growing controversy^{1,5,6,19,27}. It has been a focus of recent debates whether climate cooling at this time was caused by the final collapse of the Laurentide ice near Hudson Bay and the drainage of large glacial lakes to the North Atlantic Ocean that altered the North Atlantic thermohaline circulation. Our high-temporal-resolution data of $\delta^{18}\text{O}$ and varve thickness from Deep Lake suggest that there existed two separate, abrupt climate events between 10.0 and 8.0 cal. kyr BP. The first event, characterized by decreased $\delta^{18}\text{O}$ with an overall increase in varve thickness 9.0–8.3 cal. kyr BP, was probably related to the deglaciation in the Hudson Bay region. The second event, characterized by peak varve-thickness values at 8.2 cal. kyr BP, was probably a manifestation of an abrupt climate event of unknown cause at a global scale. These data together provide evidence for the climatic complexity of the early Holocene and for the climatic linkages between the North Atlantic region and the mid-continent of North America. □

Received 15 June 1998; accepted 17 June 1999.

- Alley, R. et al. Holocene climatic instability: A prominent, widespread event 8200 yr ago. *Geology* **25**, 483–486 (1997).
- Blunier, T., Chappellaz, J., Schwander, J., Stauffer, B. & Raynaud, D. Variations in atmospheric methane concentration during the Holocene Epoch. *Nature* **374**, 46–49 (1995).
- O'Brien, S. R. et al. Complexity of Holocene climate as reconstructed from a Greenland ice core. *Science* **270**, 1962–1964 (1995).
- Wright, H. E. Jr et al. *Global Climate Since the Last Glacial Maximum* (Univ. Minnesota Press, Minneapolis, 1993).
- von Grafenstein, U., Erlenkeuser, H., Muller, J., Jouzel, J. & Johnsen, S. The cold event 8200 years ago documented in oxygen isotope records of precipitation in Europe and Greenland. *Clim. Dyn.* **14**, 73–81 (1998).
- Klitgaard-Kristensen, D., Sejrup, H. P., Hafliadason, H., Johnsen, S. & Spurk, M. A regional 8200 cal. kyr BP cooling event in northwest Europe, induced by final stages of the Laurentide ice-sheet deglaciation? *J. Quat. Sci.* **13**, 165–169 (1998).
- Stuiver, M. & Reimer, P. J. Extended ^{14}C age calibration program. *Radiocarbon* **35**, 215–230 (1993).
- Hu, F. S., Wright, H. E. Jr, Ito, E. & Lease, K. Climatic effects of glacial Lake Agassiz in the midwestern United States during the last deglaciation. *Geology* **25**, 207–210 (1997).
- Teller, J. T. & Clayton, L. *Glacial Lake Agassiz* (Spec. Pap. 26, Geological Assoc. of Canada, Ottawa, 1983).
- Dansgaard, W. Stable isotopes in precipitation. *Tellus* **5**, 461–469 (1964).
- Rozanski, K., Araguas-Araguas, L. & Gonfiantini, R. in *Climate Change in Continental Isotopic Records* (eds Swart, P. K., Lohmann, K. C., McKenzie, J. & Savin, S.) 1–36 (American Geophysical Union, Washington DC, 1993).
- Bryson, R. A. & Hare, F. K. (eds) *Climates of North America* (Elsevier, New York, 1974).
- Simplins, W. W. Isotopic composition of precipitation in central Iowa. *J. Hydrol.* **172**, 185–207 (1995).
- Andrews, J. T. & Falconer, G. Late glacial and postglacial history and emergence of the Ottawa Islands, Hudson Bay, N.W.T.: Evidence on the deglaciation of Hudson Bay. *Can. J. Earth Sci.* **6**, 1263–1276 (1969).
- Hardy, L. La déglaciation et les épisodes lacustre et marin sur les versants de la baie de James. *Géogr. Phys. Quat.* **31**, 261–273 (1977).
- Dyke, A. S. & Prest, V. K. Late Wisconsinan and Holocene history of the Laurentide Ice Sheet. *Géogr. Phys. Quat.* **41**, 237–263 (1987).
- Kerwin, M. W. A regional stratigraphic isochron (ca. 8000 ^{14}C B.P.) from the final deglaciation of Hudson Strait. *Quat. Res.* **46**, 89–98 (1996).

18. Manabe, S. & Stouffer, R. J. Two stable equilibria of a coupled ocean-atmosphere model. *J. Clim.* **7**, 5–23 (1988).
19. de Vernal, A., Hillaire-Marcel, C., von Grafenstein, U. & Barber, D. Researchers look for links among paleoclimate events. *Eos* **78**, 245–256 (1997).
20. Kutzbach, J. E. & Webb, T. III in *Global Climates since the Last Glacial Maximum* (eds Wright, H. E. Jr et al.) 5–11 (Univ. Minnesota Press, Minneapolis, 1993).
21. Wright, H. E. Jr Patterns of Holocene climatic change in the midwestern United States. *Quat. Res.* **38**, 129–134 (1992).
22. Webb, T. III, Bartlein, P. J., Harrison, S. P. & Anderson, K. H. in *Global Climates since the Last Glacial Maximum* (eds Wright, H. E. Jr et al.) 415–467 (Univ. Minnesota Press, Minneapolis, 1993).
23. Dean, W. E. & Stuiver, M. in *Elk Lake, Minnesota, Evidence for Rapid Climate Change in the North-Central United States* (eds Bradbury, J. P. & Dean, W. E.) 163–180 (Spec. Pap. 276, Geological Soc. of America, Boulder, Colorado, 1993).
24. Laird, K. R., Fritz, S. C., Cumming, B. F. & Grimm, E. C. Early-Holocene limnological and climatic variability in the northern Great Plains. *Holocene* **8**, 275–285 (1998).
25. Anderson, R. Y. in *Elk Lake, Minnesota: Evidence for Rapid Climate Change in the North-Central United States* (eds Bradbury, J. P. & Dean, W. E.) 45–67 (Spec. Pap. 276, Geological Soc. America, Boulder, Colorado, 1993).
26. Sprowl, D. R. in *Elk Lake, Minnesota: Evidence for Rapid Climate Change in the North-Central United States* (eds Bradbury, J. P. & Dean, W. E.) 69–74 (Spec. Pap. 276, Geological Soc. of America, Boulder, Colorado, 1993).
27. Barber, D. C. et al. Forcing of the cold event of 8,200 years ago by catastrophic drainage of Laurentide lakes. *Nature* **400**, 344–348 (1999).
28. McCrea, J. M. On the isotopic chemistry of carbonates and paleo-temperature scale. *J. Chem. Phys.* **18**, 849–857 (1950).
29. Groote, P. M., Stuiver, M., White, J. W. C., Johnsen, S. & Jouzel, J. Comparison of oxygen isotope records from GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).
30. Stuiver, M., Groote, P. M. & Brazunas, T. F. The GISP2 $\delta^{18}\text{O}$ record of the past 16,500 years and the role of the Sun, Ocean and volcanoes. *Quat. Res.* **44**, 341–354 (1995).

Acknowledgements. We thank J. Teranes, B. Ammann, T. A. Brown, W. E. Dean, D. R. Engstrom and C. J. Mock for comments on the manuscript. This work was supported by the US National Science Foundation.

Correspondence and requests for materials should be addressed to E.S.H. at the Department of Plant Biology (e-mail: fshu@life.uiuc.edu).

Large-scale freshening of intermediate waters in the Pacific and Indian oceans

Annie P. S. Wong*, Nathaniel L. Bindoff† & John A. Church‡

* Institute of Antarctic and Southern Ocean Studies, University of Tasmania, GPO Box 252-77, Hobart, 7001, Australia

† Antarctic Co-operative Research Centre, GPO Box 252-80, Hobart, 7001, Australia

‡ CSIRO Division of Marine Research, GPO Box 1538, Hobart, 7001, Australia

Despite the central role of the oceans in the global hydrological cycle, direct observations of precipitation over the oceans are too sparse to infer global patterns of variability. For the regions of water-mass formation (the high latitudes), however, it is possible to obtain indirect information on changes in the surface salinity budget from salinity measurements elsewhere, as water masses in the ocean carry distinct signatures in temperature and salinity over long distances. Here we present a comparison of historical hydrographic data collected between 1930 and 1980^{1,2} with six more-recent trans-oceanic hydrographic sections (1985–94) from the intermediate waters of the Pacific and Indian oceans^{3,4}. North Pacific Intermediate Water and Antarctic Intermediate Water both show coherent basin-wide salinity decreases with time. The simplest explanation for these changes is a freshening of surface waters, over approximately 22 years, in the high-latitude North Pacific and Southern oceans, suggesting that precipitation (minus evaporation) has increased over the polar gyres. We estimate an increase by about 31 mm yr⁻¹ for the Southern Ocean (between 55° S and 65° S), which is about three times larger than the values suggested by coupled atmosphere–ocean models with increasing atmospheric greenhouse-gas concentrations for the same period^{5–8}. The patterns of change are, however, qualitatively consistent between models and observations, and our results provide evidence for an intensification of the global hydrological cycle over the past decades.

Observations of changes in the Atlantic Ocean's temperature and salinity structure^{9–12} all suggest that the circulation and water-mass characteristics of the deep ocean have significant variability. Results from the South Pacific Ocean and Tasman Sea show coherent warming on basin-scales in Subantarctic Mode Water (300–700 m depth)^{13,14}. Here we extend the geographical and depth coverage of these earlier results to North Pacific Intermediate Water (NPIW; 400–700 m) and Antarctic Intermediate Water (AAIW; 700–1,300 m), using six new zonal sections from the World Ocean Circulation Experiment (WOCE)^{3,4} in the Pacific and Indian oceans (Fig. 1). These six trans-oceanic sections are located as follows: at 47° N (1985), 24° N (1985) and 10° N (1989) respectively in the subpolar gyre, subtropical gyre and equatorial region of the North Pacific; at 17° S (1994) in the subtropical gyre of the South Pacific; at 32° S (1987) in the Indian subtropical gyre; and at 43° S (1989) in the Tasman Sea¹³.

Five of the sections show a well-defined salinity minimum of either NPIW (24° N) or AAIW (10° N, 17° S, 32° S, 43° S) (Fig. 2). The shape of the temperature–salinity curve is defined by the wintertime sea surface temperature and salinity where these waters sink into the ocean interior. Thus anomalies in wintertime sea surface temperature or salinity are reflected directly as anomalies in the temperature or salinity properties of the water after subduction. (We note that this is an over-simplification of the ventilation process because the mixed layer and its temperature–salinity characteristics are also sensitive to anomalies in the winds and mixing processes within the ocean interior^{15–17}. However, this simplified model of ventilation does provide a guide to the interpretation of the results presented here.)

The differences between the historical and WOCE observations are clearly seen in the zonally averaged temperature–salinity curve for each section (Fig. 2) and are statistically significant at the 90% confidence level (Fig. 3). Section plots (not shown) of temperature and salinity differences on density surfaces all show coherent basin-wide changes as reflected in the averaged temperature–salinity curves. Along the five sections at 24° N, 10° N, 17° S, 32° S and 43° S, the salinity minimum has freshened (and cooled) on density surfaces, with the temperature–salinity curve now displaced to lower salinity values. Above the salinity minimum (that is, in the shallower layers), the displacement of the temperature–salinity curve to cooler/fresher values can be explained by a combination of warming and freshening of surface waters¹⁸. However, at the salinity minimum, changes in the wintertime sea surface temperature cannot displace the temperature–salinity curve to fresher values. Thus the observed freshening of the salinity minimum is explained most simply by a freshening of the surface waters in the source regions of NPIW and AAIW.

NPIW (indicated by the low-salinity tongue north of 15° N in Fig. 3) is confined to the North Pacific, and is observed only in the 24° N section. But along 47° N, significant freshening on density surfaces in the NPIW range (density surfaces 26.4–27.2 kg m⁻³) is also present. The largest decreases are 0.1 practical salinity units (p.s.u.) along 47° N, and 0.017 p.s.u. along 24° N (Fig. 3a). For the 24° N section, the NPIW density surfaces have been displaced downwards on average by 22 m (Fig. 3b) and thickened. This thickening of the NPIW layer along 24° N has led to an increase in the vertical cross-sectional area of this water mass by 1.4%. The largest difference occurs in the 47° N section, with a weaker freshening signal in the 24° N section, consistent with the increased distance of 24° N from the source of NPIW in the Kuril Islands/Kamchatka regions and in the Sea of Okhotsk¹⁹.

AAIW is found in all of the Southern Hemisphere oceans and parts of the North Atlantic and North Pacific oceans. AAIW (indicated by the low-salinity tongue south of the Equator in Fig. 3, and density surfaces 27.2–27.6 kg m⁻³) is observed along the four sections at 10° N, 17° S, 32° S and 43° S, and shows significant freshening on density surfaces. The largest differences

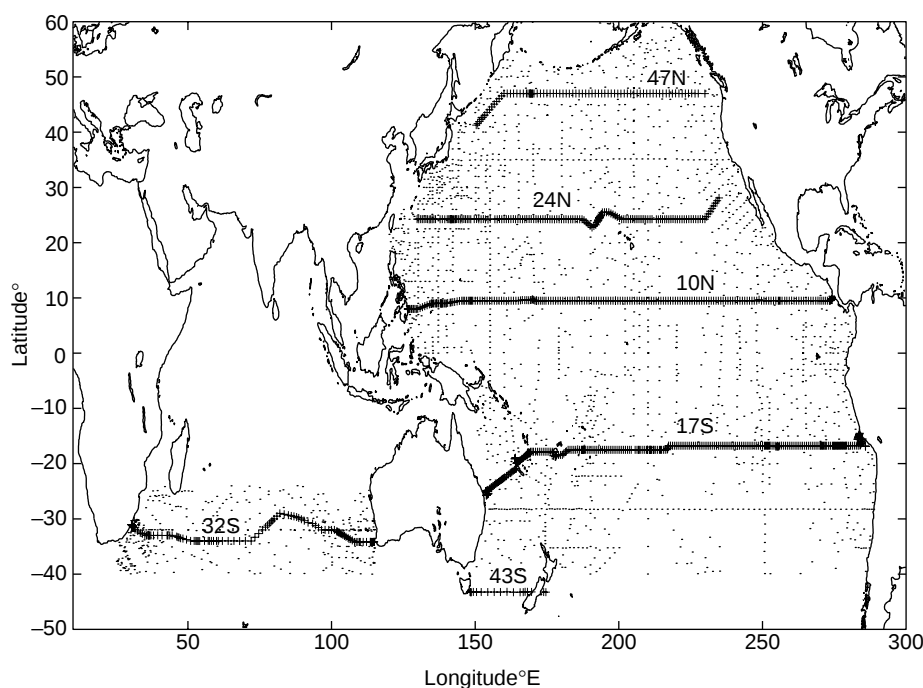


Figure 1 Locations of data used in this study. The selected historical data are shown by dots, and the modern WOCE data are shown by crosses. The WOCE

sections are designated 47°N, 24°N, 10°N, 17°S, 32°S and 43°S. The selected historical data are available²⁵.

in AAIW occur in the 17°S and 32°S sections (respectively 0.021 and 0.064 p.s.u.). The 17°S section shows no significant displacements of the density surfaces (Fig. 3b). The sections 10°N, 32°S and 43°S all show the same pattern of significant downward displacement of density surfaces, with a thinning of the AAIW layer and a reduction of the area of this water mass in each section by respectively 3.9%, 4.8% and 1.8%. The reduction in sectional area suggests a decrease in the subducted volume of water, and the non-uniform depth differences between the sections in AAIW (density range 27.2–27.6 kg m⁻³) suggest small changes in the circulation.

The deeper water masses, represented by density surfaces greater than 27.8 kg m⁻³ (at depths typically greater than 1,600 m), show no significant differences over the measurement period of 22 years (at the 90% confidence level) in the zonal averages of salinity on density surfaces, and are therefore not shown in Fig. 3. This preservation of the deep-water characteristics is consistent with the smoothing of short-term variability of subducted source waters by mixing along the circulation path¹⁷, and the long timescales (compared with the observational period) required for the transport of these waters from their sources in the Southern and North Atlantic oceans.

In the South Pacific, the main region where polar surface waters enter the ocean interior as AAIW is in the subantarctic zone in the southeastern Pacific²⁰. It then flows north around the South Pacific subtropical gyre and across the Equator in the western Pacific, or through the Drake Passage as part of the Antarctic Circumpolar Current and into the Atlantic and Indian oceans. This circulation pattern of AAIW is consistent with the 17°S section having a stronger freshening signal than at 10°N. The relatively fast Antarctic Circumpolar Current spreads AAIW in the Indian Ocean sector, and hence across the 32°S section in the Indian Ocean, plausibly explaining the strong freshening observed there. In the Tasman Sea, dynamic topography maps²¹ show that the 43°S section is fed from the south by the Antarctic Circumpolar Current and from the north by the South Pacific subtropical gyre. For both of these paths, the 43°S section is more distant from the sources of AAIW than the Indian Ocean sector and so is consistent with the weaker freshening signal.

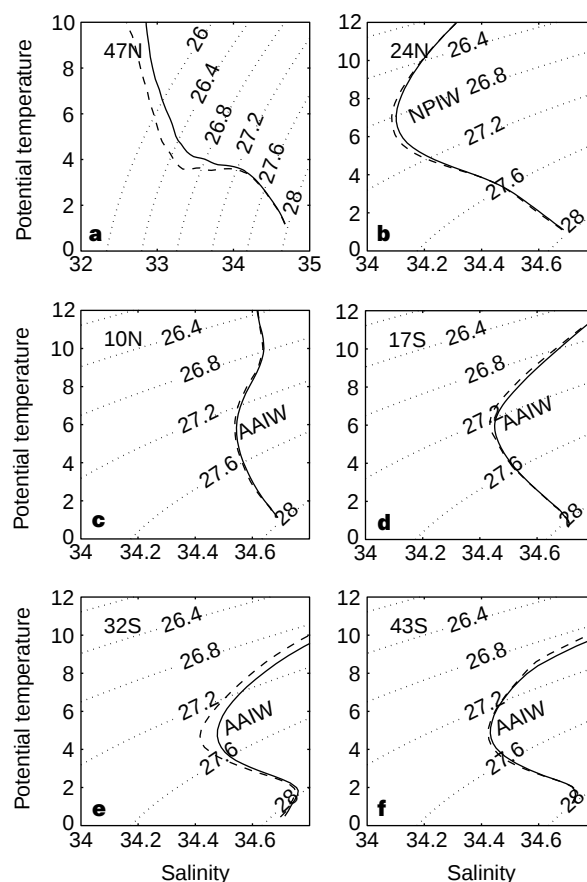


Figure 2 Zonally averaged temperature-salinity curves. Data are shown along each of the six sections: **a, b, c, d, e** and **f** are respectively for 47°N, 24°N, 10°N, 17°S, 32°S and 43°S. The thick dashed and solid lines are for the modern and historical data, respectively. Lines of constant density (dotted) are superimposed. Note that the scales in **a** are different to the other panels.

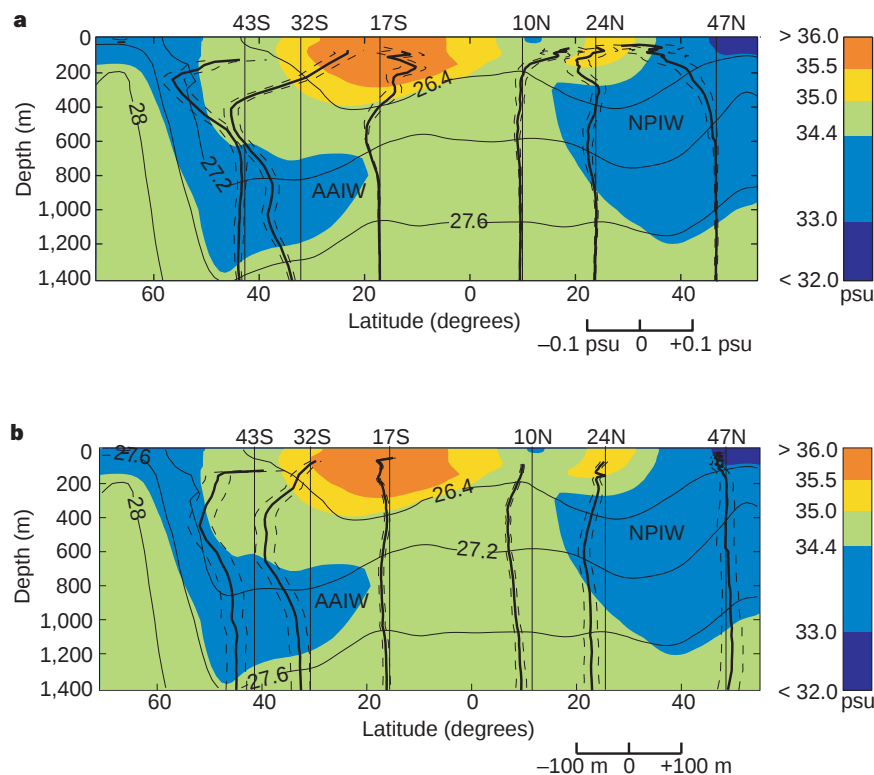


Figure 3 Zonally averaged differences on density surfaces. **a**, Differences of salinity, and **b**, differences of depth, for each of the six sections at 47°N, 24°N, 10°N, 17°S, 32°S and 43°S. The differences on density surfaces (solid lines) are plotted on depth coordinates. The error margins (dashed lines) are the 90% confidence intervals based on *t*-test criteria and take into account the degrees of freedom, the error in differencing the two fields, and the natural variability of the

salinity (or depth) field caused by mesoscale eddies and seasonal variations (but not inter-annual variability). The differences are superimposed over a background of the mean salinity field along 200°E in the Pacific Ocean³⁰. The approximate depth positions of the 26.4, 27.2 and 27.6 kg m⁻³ density surfaces are also shown (contours). AAIW is bounded by density surfaces 27.2 and 27.6 kg m⁻³, and NPIW is bounded by density surfaces 26.4 and 27.2 kg m⁻³.

Extrapolating the mean observed salinity decrease for AAIW to the entire South Pacific (mean salinity difference of 0.02 p.s.u. over a mean thickness of 500 m), the freshwater gain in the source region of AAIW is estimated at 31 mm yr⁻¹. This is ~8.5% of the annual rainfall over the ocean south of 55°S. Unfortunately, direct observations of rainfall (and salinity) over the Southern Ocean are sparse and unreliable. However, results from ice cores from Wilkes Land, Antarctica, show that snow accumulation has increased over a similar period by 23% (ref. 22). In the source region of AAIW, coupled ocean-atmosphere models (using observed greenhouse-gas increases for the same period as the observations) show increasing freshwater input of 8 mm yr⁻¹ (ref. 5) and 12.5 mm yr⁻¹ (ref. 6). Although our observation-based estimate of 31 mm yr⁻¹ is about 3 times larger than the model estimates, the spatial pattern of freshening inferred from the observations in both the North Pacific and the Southern Oceans is qualitatively consistent with these models.

The observed differences in the ocean could also be decadal variability in high-latitude precipitation. However, there must be a net increase in precipitation (minus evaporation) in both hemispheres to explain the coherent freshening signal in both NPIW and AAIW. The Antarctic Circumpolar Wave²³, the observed decrease in the strength of the Circumpolar Trough and changes in the semi-annual oscillation²⁴ could also have important roles in this variability.

The three Pacific sections at 24°N, 10°N and 17°S show a net decrease in salinity over the water column (excluding the mixed layer and the deeper waters where salinity differences are unreliable) of respectively 0.0049 p.s.u., 0.0016 p.s.u. and 0.0068 p.s.u. These relatively small decreases in depth-integrated salinity result from a combination of decreased salinity in the intermediate waters and

increased salinity in the shallower thermocline waters that originate from the high-evaporation regions of the subtropical gyres (Fig. 3a, 10°N, 17°S, 32°S). In these sections, the increase in salinity above the salinity minimum balances 29–60% of the observed freshening of AAIW and NPIW. The combined changes suggest a strengthening of the hydrological cycle over the past 22 years. These results and other similar observations^{9–12,14} have an important role in validating and testing coupled ocean-atmosphere models used for predictions of climate change and climate variability. □

Methods

For comparison with the WOCE sections, a total of 2,383 hydrographic profiles of observations of the full ocean depth from the Pacific and Indian oceans (Fig. 1) were selected from two historical hydrographic atlases^{1,2}. These NODC data were selected on the basis of geographical coverage and quality (ref. 25 and J. Reid, personal communication). In addition, we further screened each vertical profile of temperature and salinity for outliers relative to near neighbours. The final set of selected historical data includes profiles from 1930 to 1980. However, nearly all of the historical data used were collected over

Table 1 Modern and historical data used in the comparison						
Zonal section	Modern (year)	No. of casts	Historical (year)	No. of casts	Standard deviation (years)	Time difference (years)
47°N	1985	299	1966	299	5.0	19
24°N	1985	156	1970	509	5.5	15
10°N	1989	213	1969	400	6.0	20
17°N	1994	294	1967	706	5.0	30
32°S	1987	109	1962	451	10.0	25
43°S	1989	72	1967	18	<1	22

a relatively narrow time period in the 1960s as part of the International Geophysical Year programs (Table 1). Thus, for interpretation, the average time difference between the historical and WOCE data is 22 years with an uncertainty of 6 years.

To compare the historical and WOCE-period data, the former were interpolated onto the geographical coordinates of the latter. The 47°N, 24°N, 10°N and 17°S sections were interpolated on neutral density surfaces²⁶, and the earlier results at 32°S and 43°S sections were interpolated on standard depth surfaces using empirical objective functions and objective mapping techniques^{13,27}. Although interpolating on pressure surfaces is less accurate²⁸ in the case of the 32°S and 43°S sections, the errors are less than instrumental precision. The main advantages of using salinity and temperature on density surfaces are that they are less sensitive to short-term oceanographic processes and can be more readily interpreted in terms of changed sea surface temperature and salinity^{10,13,18,29}.

Received 15 February; accepted 8 June 1999.

- Gordon, A., Molinelli, E. J. & Baker, T. N. *Southern Ocean Atlas* (Alfred Wegener Inst. Amerind, New Delhi, 1982).
- World Ocean Atlas 1994: CD-ROM data set (Nat'l Oceanographic Data Center, Ocean Climate Lab., Washington DC, 1994).
- WOCE Data Products Committee *WOCE Global Data, Version 1.0* CD-ROM; WOCE Rep. 158/98 CD-ROM (WOCE Int. Project Office, Southampton, UK, 1998).
- WOCE Hydrographic Program Office *WOCE Hydrographic Program Data* (Scripps Inst. of Oceanography, San Diego, 1998); also available at (<http://whpo.ucsd.edu>) (cited 31 Aug. 1998).
- Manabe, S., Bryan, K. & Spelman, M. J. Transient response of a global ocean-atmosphere model to a doubling of atmospheric carbon dioxide. *J. Phys. Oceanogr.* **20**, 722–749 (1990).
- Gordon, H. B. & O'Farrell, S. P. Transient climate change in the CSIRO coupled model with dynamic sea ice. *Mon. Weath. Rev.* **125**, 875–907 (1997).
- Manabe, S., Stouffer, R. J., Spelman, M. J. & Bryan, K. Transient responses of a coupled ocean-atmosphere model to gradual changes of atmospheric CO₂. Part I: annual mean response. *J. Clim.* **4**, 785–818 (1991).
- Murphy, J. M. & Mitchell, J. F. B. Transient response of the Hadley Centre coupled ocean-atmosphere model to increasing carbon dioxide. Part II: Spatial and temporal structure of response. *J. Clim.* **8**, 57–80 (1995).
- Levitus, S. Interpentadal variability of temperature and salinity at intermediate depths of the North Atlantic Ocean, 1970–1974 versus 1955–1959. *J. Geophys. Res.* **94**, 6091–6131 (1989).
- Bryden, H. L. *et al.* Decadal changes in water mass characteristics at 24N in the subtropical North Atlantic Ocean. *J. Clim.* **9**, 3162–3186 (1996).
- Dickson, R., Lazier, J., Meincke, J., Rhines, P. & Swift, J. Long-term coordinated changes in the convective activity of the North Atlantic. *Prog. Oceanogr.* **38**, 241–295 (1996).
- Read, J. F. & Gould, W. J. Cooling and freshening of the subpolar North Atlantic Ocean since the 1960s. *Nature* **360**, 55–57 (1992).
- Bindoff, N. L. & Church, J. A. Warming of the water column in the southwest Pacific Ocean. *Nature* **357**, 59–62 (1992).
- Johnson, G. C. & Orsi, A. H. Southwest Pacific Ocean water mass changes between 1968/9 and 1990/1. *J. Clim.* **10**, 306–316 (1997).
- Joyce, T. M., Luyten, J. R., Kubryakov, A., Bahr, F. B. & Pallant, J. S. Meso- to large-scale structure of subducting water in the subtropical gyre of the eastern North Atlantic Ocean. *J. Phys. Oceanogr.* **28**, 40–61 (1998).
- Hazeleger, W. & Drijfhout, S. S. Mode water variability in a model of the subtropical gyre: response to anomalous forcing. *J. Phys. Oceanogr.* **28**, 266–288 (1998).
- Hazeleger, W. & Drijfhout, S. S. Stochastically forced Mode Water variability. *J. Phys. Oceanogr.* (in the press).
- Bindoff, N. L. & McDougall, T. J. Diagnosing climate change and ocean ventilation using hydrographic data. *J. Phys. Oceanogr.* **24**, 1137–1152 (1994).
- Warner, M. J., Bullister, J. L., Wisegarver, D. P., Gammon, R. H. & Wiess, R. F. Basin-wide distributions of chlorofluorocarbons CFC-11 and CFC-12 in the North Pacific: 1985–1989. *J. Geophys. Res.* **101**, 20525–20542 (1996).
- McCartney, M. S. in *A Voyage of Discovery* (ed. Angel, M.) 103–119 (Pergamon, Oxford, 1977).
- Reid, J. L. On the total geostrophic circulation of the South Pacific Ocean: flow patterns, tracers and transports. *Prog. Oceanogr.* **16**, 1–61 (1986).
- Morgan, V. I., Goodwin, I. D., Etheridge, D. M. & Wooley, C. W. Evidence from Antarctic ice cores for recent increases in snow accumulation. *Nature* **354**, 58–60 (1991).
- White, W. B. & Peterson, R. G. An Antarctic circumpolar wave in surface pressure, wind, temperature, and sea-ice extent. *Nature* **380**, 699–702 (1996).
- Hurrell, J. W. & Loon, H. V. A modulation of the atmospheric annual cycle in the Southern Hemisphere. *Tellus A* **46**, 325–338 (1994).
- Reid, J. & Mantyla, A. *World Dataset* (cited Jan. 1999) (<ftp://minerva.ucsd.edu>) (1995).
- Jackett, D. & McDougall, T. J. A neutral density variable for the world's oceans. *J. Phys. Oceanogr.* **27**, 237–263 (1997).
- Bindoff, N. L. & Wunsch, C. Comparison of synoptic and climatologically mapped sections in the South Pacific Ocean. *J. Clim.* **5**, 631–645 (1992).
- Lozier, M. S., McCartney, M. S. & Owens, W. B. Anomalous anomalies in averaged hydrographic data. *J. Phys. Oceanogr.* **24**, 2624–2638 (1994).
- Church, J. A., Godfrey, J. S., Jackett, D. R. & McDougall, T. J. A model of sea level rise caused by ocean thermal expansion. *J. Clim.* **4**, 438–456 (1991).
- Levitus, S. *Climatological Atlas of the World Ocean* (Nat'l Oceanographic and Atmospheric Administration, US Dept of Commerce, Washington DC, 1982).

Acknowledgements. We thank H. Bryden for access to the WOCE P21 section at 17°S before publication, and R. Dickson for discussions about changes in the semi-annual oscillation. A.W. was supported by a National Greenhouse Advisory Committee scholarship. This Letter is a contribution to the World Ocean Circulation Experiment and the CSIRO Climate Change Research Program.

Correspondence and requests for materials should be addressed to N.L.B. (e-mail: n.bindoff@utas.edu.au).

Rigidity variations with depth along interplate megathrust faults in subduction zones

Susan L. Bilek & Thorne Lay

Institute of Tectonics and Earth Sciences Department, University of California, Santa Cruz, California 95064, USA

The world's largest earthquakes occur along the contact between subducting and overriding tectonic plates in subduction zones¹. Rock and sediment properties near this plate interface exert important controls on the frictional behaviour of faults and earthquake rupture dynamics². An important material property

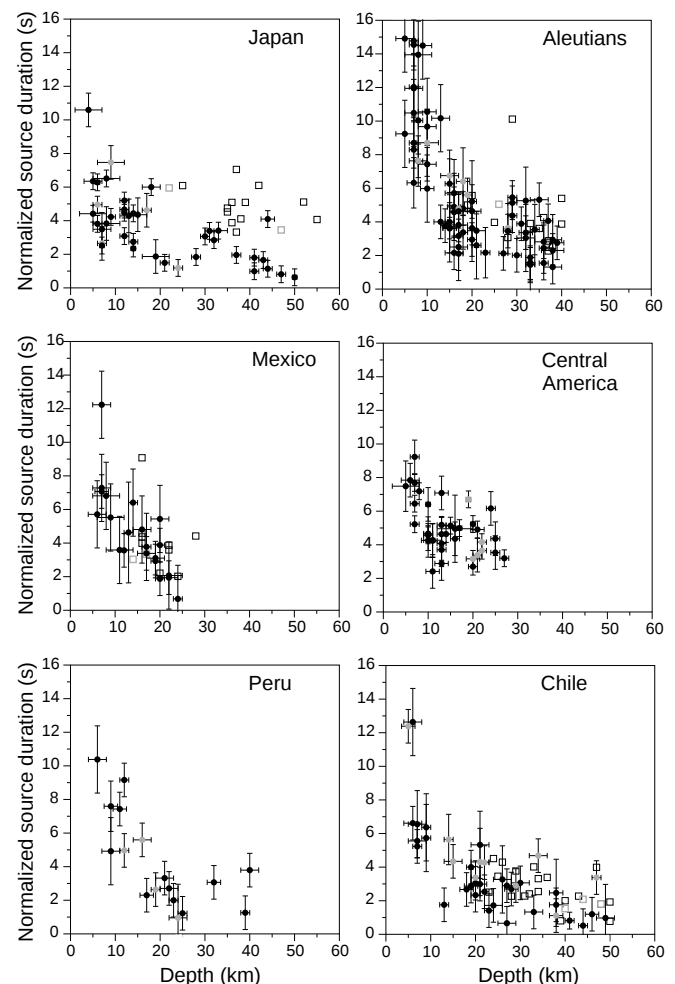


Figure 1 Normalized source duration as a function of source depth (depth below ocean bottom) for each of the regions analysed. All of the 291 event source durations were scaled using the cube root of seismic moment, normalized to a $M_w = 6.0$ event (M_0 of 1.16×10^{25} dyne cm). Open symbols indicate the 61 events analysed in a similar fashion from an earlier study of maximum coupling depth^{9,28}. Grey circles represent thrust events satisfying our initial event criteria that have depth estimates indicating that they do not lie along the plate interface, but are instead events in the accretionary wedge or intra-plate events. The overall trend of decreasing source duration with increasing depth is consistent between regions. Differences in ranges of source duration and depth may be related to the thickness of sediment in the trench (for example, the Aleutians have on average the most sediment in the trench, while Central America has the least^{17–19}) and to other factors discussed in the text.

to define along the plate interface is the rigidity (the resistance to shear deformation). Rigidity affects the degree of earthquake shaking generated by a given fault displacement through its influences on seismic wave speed and earthquake rupture velocity. Here we present an investigation of the relationship between the duration of earthquake rupture and source depth, which yields estimates of rigidity variation along plate interfaces in six subduction zones in the circum-Pacific region. If stress drop is assumed constant, rigidity appears to increase with depth in each seismogenic zone by a factor of ~ 5 between depths of 5 and 20 km. This result is consistent with the hypothesis that 'tsunami' earthquakes (characterized by large slip for a given seismic moment and slow rupture velocity) occur in regions of low rigidity at shallow depths³⁻⁵. These rigidity trends should provide an important constraint for future fault-zone and earthquake-modelling efforts.

We analysed hundreds of subduction zone events from northern Japan⁶, the Alaska–Aleutian islands region, Mexico, Central America⁷, Peru and Chile satisfying the following criteria: (1) close proximity to the interplate contact, (2) faulting geometry consistent with interplate thrusting, (3) moment magnitude (M_w) between 5.0 and 7.5, and (4) availability of at least four high-quality teleseismic digital P-wave recordings with good azimuthal coverage and high signal to noise ratio. (Event data are available; see Supplementary Information.) For each event, we obtained source time functions by using, on average, 10 broadband P-wave records in a multi-station point-source deconvolution method⁸⁻¹⁰ using a half-space model with a P-wave velocity of 6.0 km s^{-1} for a range of trial depths. The optimal source depth is identified by minimizing the misfit between the data and synthetic seismograms, and the source duration is measured from the corresponding source time function, which is flexibly parametrized as a time series of boxcars

with variable amplitude. Typically, the optimal source depth estimate has an uncertainty of several kilometres over which we obtain comparable low misfit values, and we choose the depth with the 'simplest' rupture, with energy release concentrated early in the source function^{9,11}. The source duration is measured for the primary pulse of the source time function, which is usually sufficient to characterize the rupture. Identifying the termination point for the primary pulse is sometimes problematic due to the limited bandwidth of the data and errors in the model Green's functions. We strive to be consistent in measuring the durations as the time extent of the primary pulse, ending when the source time function returns to a baseline level, and include error bars on our measurements.

Because rupture duration is expected to increase with the cube root of seismic moment (M_0 ; ref. 12), we scale measured durations to a common earthquake size. We divide each measured duration by the cube root of the event seismic moment from the Harvard centroid moment tensor (CMT) catalogue, scaled to an event of $M_w = 6$, a procedure similar to that used by others¹³⁻¹⁵. This moment-scaling removes the trend of increasing duration with increasing moment. The source function duration also depends on details of the rupture geometry (for example, bilateral, unilateral, circular, irregular) which are not known and which cannot be reliably measured from the available data. We therefore assume that a simple unilateral rupture model applies to all our events. Using a circular rupture model gives similar results.

All six regions show robust trends of decreasing normalized source duration with increasing depth below the sea floor (Fig. 1). Comparison of our duration estimates to prior work of a similar nature¹³⁻¹⁵ indicates compatibility after allowing for differences in the reference moment used in the scaling (Fig. 1 includes some corresponding measurements). The subjectivity of determining time function termination can account for some differences in

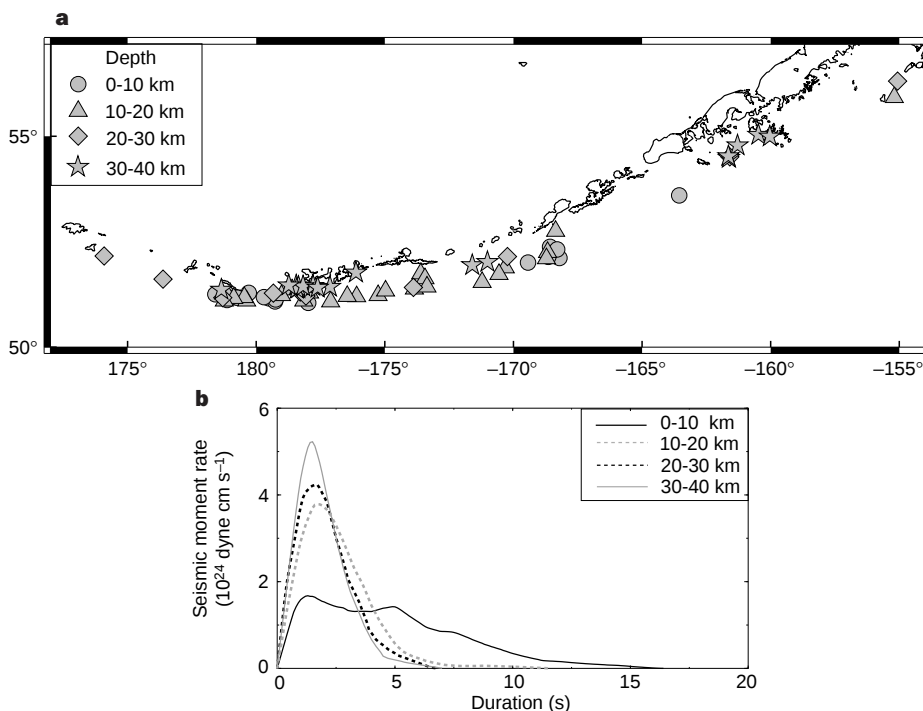


Figure 2 Data from the Alaska–Aleutian islands region. **a**, Map showing event locations from the National Earthquake Information Center (NEIC) as well as the optimal depths of the earthquakes, as determined by our inversion. **b**, Average normalized source time functions. The seismic moment and duration of the time functions were both scaled to remove the effects of moment¹⁵, before summing in depth bins. The timescales were moment-scaled. The moment rate axes were divided by $M_0^{2/3}$ in order to maintain the relationship between moment and area under the time function. The time function was set to 0 before the beginning of the

large energy pulse of the source time function, and 0 after the time we chose for the ending of the energy pulse. The source functions have been scaled such that the area beneath each curve is equal to $1.16 \times 10^{26} \text{ dyne cm}$, the moment of the reference event of $M_w = 6.0$. Bin 0–10 km contains 20 events, bin 10–20 km contains 26 events, bin 20–30 km contains 10 events, and bin 30–40 km contains 18 events. Events at shallow depth ($< 10 \text{ km}$) have by far the longest average time function (16.4 s), while events at depths $> 30 \text{ km}$ have the shortest average time function (7.0 s).

duration estimates. There are minor variations in the relationship between regions, but we concentrate on the common trend. The Alaska–Aleutian data set (Fig. 2) best illustrates the dramatic depth dependence of source duration behaviour.

Given the overall consistency of the rupture duration–depth relationship for all six regions, it is viable to characterize depth-dependent material properties of the subduction-zone interfaces. Recognizing that rupture duration is influenced by many factors, we use the observed source duration variations to constrain an end-member model with constant stress drop, involving variations in rigidity. Rigidity (μ), a measure of the proportionality between shear stress and shear strain, influences source duration through an empirical dependence of rupture velocity on shear wave velocity. The standard definitions of seismic moment (M_0) and stress drop ($\Delta\sigma$) are

$$M_0 = \mu S D \propto \Delta\sigma w^2 l \approx \Delta\sigma L^3 \quad (1)$$

$$\Delta\sigma = C\mu(D/L) \quad (2)$$

where S is fault area, l and w are the fault length and width, with general dimension L , D is average displacement, and C is a constant for a given fault geometry¹². Using these definitions, along with the empirical observation that earthquake rupture velocity (V_r) is typically about $80\% \pm 10\%$ of the shear wave velocity (β), and the fact that source duration (τ) is related to the rupture velocity (V_r) for a simple unilateral rupture by

$$\tau = L/V_r \approx L/(0.8\beta) = L/(0.8\sqrt{\mu/\rho}) \quad (3)$$

we can estimate rigidity using τ . In doing so, we assume constant stress drop, a source dimension, L , of 10 km, and a constant density, ρ , of 2.7 g cm^{-3} , assumptions that are reasonable for a $M_w = 6.0$ event within the volumetric seismogenic zone. Rigidity can now be estimated by:

$$\mu \approx \rho L^2 / (0.8^2 \tau^2) \quad (4)$$

The choice of source dimension and density affects the baseline of the rigidity values, but does not change the relative magnitude of depth variations. Although there is undoubtedly some scatter due to actual density variations, differences in rupture mode from event to event, and noise in the measurements, we proceed to convert our scaled durations to estimates of rigidity.

The estimated rigidity of the seismogenic zone increases over the depth range 5–50 km when we consider our entire data set (Fig. 3a). There is large scatter in the data, but the overall trend is clear. At depths between 20 and 40 km, our average rigidity values are very similar to those for the Preliminary Reference Earth Model (PREM), indicating sensible baseline parameters. Each region shows a similar increase in rigidity with increasing depth. These data constitute, to our knowledge, the first systematic measure of how an important material property changes with depth along the subduction-zone interface. A corresponding increase of rupture velocity with increasing source depth is implied, along with decreasing magnitude of slip for a given seismic moment. A concern is that the Harvard CMT moments used to scale the durations are calculated using a PREM rigidity. If the rigidity does indeed vary as in Fig. 3a, the CMT seismic moments will be too small for the shallow events, and thus their scaled durations may be too long. Correcting for this quantitatively is very difficult because the rigidity decrease is probably in a three-dimensional wedge, and numerical methods are needed to calculate seismic wave excitation. However, the trend would be slightly reduced. The source depths may also be slightly overestimated, because the shallow rock velocities are lower than our half-space model predicts.

Mapping the source duration variations into rigidity variations with depth is one end-member model that can explain the observations. Another end member is the case of variable stress drop assuming a constant rigidity. All events then have equal rupture velocity, but their rupture area varies systematically. The resulting

stress-drop estimates show increasing stress drop with increasing depth (Fig. 3b). To resolve the trade-off between fault area and rupture velocity, we need an independent way to determine the fault area for each event, but this is not currently possible. It is very likely that strong variations in rigidity may couple with stress-drop variations and that the actual situation is somewhere in between the two end members, but there are independent reasons to expect low rigidities at shallow depths.

In many subduction zones, drilling, geochemical and seismic surveys suggest that sediment is being subducted beneath the overriding plate^{16–21}. Increasing compaction and fluid expulsion alter the material with depth along the seismogenic zone, and a variety of mineralogical phase changes within the subducting sediments, and in the subducting plate, are expected to occur with increasing pressure and temperature (Fig. 4). The smectite to

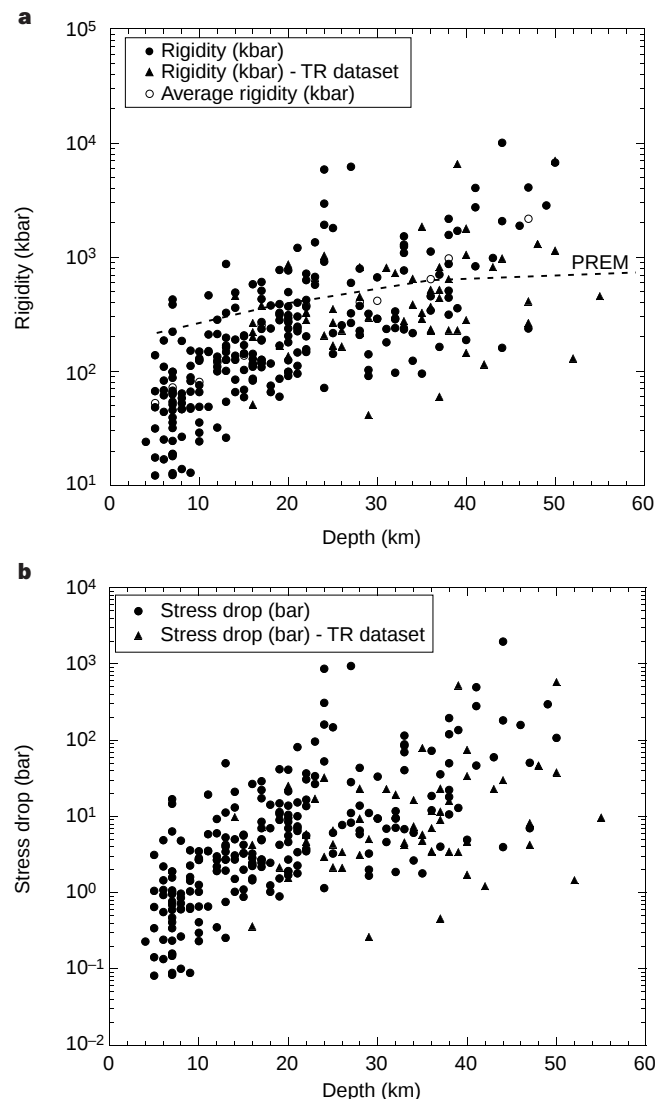


Figure 3 Two end-member models for our data. **a**, Plot of estimated rigidity variations along the megathrust with depth for the entire 291 events in the data set. Filled circles indicate rigidity values for individual events, triangles indicate rigidity estimates for events of the Tichelaar and Ruff (TR) data set²⁸, and open circles are average rigidity values in 2-km intervals. The dashed line shows rigidity variations estimated from PREM shear wave velocity and density values²⁹. At shallow depths, the PREM model is given at only a few depths. Our rigidity estimates involve a volumetric average rigidity within the high-strain seismogenic zone, and do not apply to the intraplate environment represented by PREM. **b**, Plot of stress-drop variations along the megathrust with depth. L is derived from τ using a constant V_r of 3.5 km s^{-1} , and $\Delta\sigma$ is calculated from equation (1).

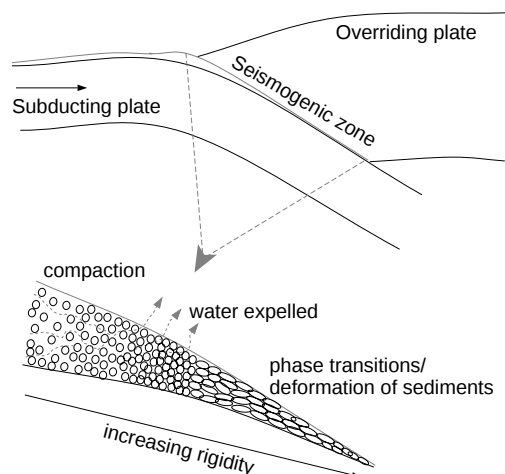


Figure 4 Generalized diagram of a subduction zone, with an expanded-scale view of the seismogenic-zone interface. As the high-porosity and un lithified sediments are subducted, they undergo several changes that can alter their rigidity. Increased pressure with depth causes compaction of the subducting sediments, expulsion of fluids, metamorphism, and mineralogical phase changes, resulting in higher rigidity of the material with increasing depth. Using our data to look at the rigidity variation end-member model, we have begun to quantify the magnitude of the changes of mechanical properties that occur with depth in the subduction zone.

illite transition, which occurs at temperatures of approximately 100–110 °C and depths of a few kilometres, depending on subduction convergence rates²², is one such phase transition expected in the subducted sediments. The change from smectite, a weaker hydrated clay, to illite, a much stronger, dehydrated clay could lead to increased rigidities with increasing depths. Another phase transition occurring within subducting oceanic crust involves metamorphosis of basalt to eclogite, with an expulsion of fluid during the breakdown of water-bearing clays and amphiboles²³. These phase transitions may play a role in changing the overall properties of the subduction-zone interface; however, our rigidity estimates involve a volumetric averaging of material in the seismogenic zone, so a variety of different mechanisms may contribute to the pattern in Fig. 3.

The model of rigidity variation with depth is particularly important in light of the occurrence of tsunami earthquakes, which produce unusually strong tsunamis for their seismic moment³. Several mechanisms have been proposed to account for tsunami earthquakes, including submarine landslides and earthquake rupture through low-rigidity sediments, either along the interface between the plates or within the forearc prism above the interface. For those events for which submarine landslides have been ruled out as the cause, earthquake source studies have shown that the rupture tends to occur at very shallow depths near the otherwise aseismic forearc wedge, where low-rigidity sediments are likely to be concentrated²⁴. Tsunami earthquakes, such as the 1992 Nicaragua event, have been found to typically have anomalously slow rupture velocities, suggesting that low-rigidity material controls the rupture²⁵. Reconciling the displacements needed to excite the Nicaragua tsunami with the observed seismic moment leads to rigidities 3–4 times lower than in PREM²⁶ and very close to those shown in Fig. 3. Similar modelling of the 1996 Peru tsunami also resulted in a smaller-than-average rigidity value to fit the tsunami run-up data²⁷. These tsunami events can be seen as the natural consequences of ubiquitous depth dependence of rigidity in seismogenic zones. Our results indicate that for the constant-stress-drop model, analysis of rigidity variations in different zones based on small events can guide assessment of potential for tsunami earthquakes, along with improving the accuracy of slip estimates made for interplate faulting which underlie most earthquake probability calculations. □

Received 16 February; accepted 22 June 1999.

- Kanamori, H. Rupture process of subduction-zone earthquakes. *Annu. Rev. Earth Planet. Sci.* **14**, 293–322 (1986).
- Scholtz, C. *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, 1990).
- Kanamori, H. Mechanism of tsunami earthquakes. *Phys. Earth Planet. Inter.* **6**, 246–259 (1972).
- Fukao, Y. Tsunami earthquakes and subduction processes near deep sea trenches. *J. Geophys. Res.* **84**, 2303–2314 (1979).
- Pelayo, A. M. & Wiens, D. A. Tsunami earthquakes: slow thrust faulting events in the accretionary wedge. *J. Geophys. Res.* **97**, 15321–15337 (1992).
- Bilek, S. L. & Lay, T. Variation of interplate fault zone properties with depth in the Japan subduction zone. *Science* **281**, 1175–1178 (1998).
- Bilek, S. L. & Lay, T. Comparison of depth dependent fault zone properties in the Japan trench and Middle America trench. *Pure Appl. Geophys.* (in press).
- Ruff, L. J. Multi-trace deconvolution with unknown scale factors: omnilinear inversion of P and S waves for source time functions. *Geophys. Res. Lett.* **16**, 1043–1046 (1989).
- Tichelaar, B. W. & Ruff, L. J. Seismic coupling along the Chilean subduction zone. *J. Geophys. Res.* **96**, 11997–12022 (1991).
- Ruff, L. J. & Miller, A. D. Rupture process of large earthquakes in the northern Mexico subduction zone. *Pure Appl. Geophys.* **142**, 101–171 (1994).
- Christensen, D. H. & Ruff, L. J. Analysis of the trade-off between hypocentral depth and source time function. *Bull. Seismol. Soc. Am.* **75**, 1637–1656 (1985).
- Kanamori, H. & Anderson, D. L. Theoretical basis of some empirical relations in seismology. *Bull. Seismol. Soc. Am.* **65**, 1073–1095 (1975).
- Vidale, J. E. & Houston, H. The depth dependence of earthquake duration and implications for rupture mechanics. *Nature* **355**, 45–47 (1993).
- Ekstrom, G. & Engdahl, E. R. Earthquake source parameters and stress distribution in the Adak Island region of the central Aleutian Islands, Alaska. *J. Geophys. Res.* **94**, 15499–15519 (1989).
- Houston, H., Benz, H. M. & Vidale, J. E. Time functions of deep earthquakes from broadband and short-period stacks. *J. Geophys. Res.* **103**, 29895–29913 (1998).
- Plank, T. & Langmuir, C. H. The chemical composition of subducting sediment and its consequences for the crust and mantle. *Chem. Geol.* **145**, 325–394 (1998).
- von Huene, R. & Scholl, D. W. Observations at convergent margins concerning sediment subduction, subduction erosion, and the growth of continental crust. *Rev. Geophys.* **29**, 279–316 (1991).
- Tera, F. et al. Sediment incorporation in island-arc magmas: Inferences from ¹⁰Be. *Geochim. Cosmochim. Acta* **50**, 535–550 (1986).
- Kimura, G. et al. Proc. ODP Init. Rep. Costa Rica accretionary wedge. **170**, 1–458 (1997).
- von Huene, R. & Lallemand, S. Tectonic erosion along the Japan and Peru convergent margins. *Geol. Soc. Am. Bull.* **102**, 704–720 (1990).
- Suyehiro, K. & Nishizawa, A. Crustal structure and seismicity beneath the forearc off northwestern Japan. *J. Geophys. Res.* **99**, 22331–22347 (1994).
- Vrolijk, P. On the mechanical role of smectite in subduction zones. *Geology* **18**, 703–707 (1990).
- Peacock, S. M. The importance of blueschist–eclogite dehydration reactions in subducting oceanic crust. *Geol. Soc. Am. Bull.* **105**, 684–694 (1993).
- Ihmle, P. F., Gomez, J.-M., Heinrich, P. & Guibourg, S. The 1996 Peru tsunamigenic earthquake: broadband source process. *Geophys. Res. Lett.* **25**, 2691–2694 (1998).
- Kanamori, H. & Kikuchi, M. The 1992 Nicaragua earthquake: a slow tsunami earthquake associated with subducted sediments. *Nature* **361**, 714–716 (1993).
- Satake, K. Mechanism of the 1992 Nicaragua tsunami earthquake. *Geophys. Res. Lett.* **21**, 2519–2522 (1994).
- Heinrich, P., Schindeler, F., Guibourg, S. & Ihmlé, P. M. Modeling of the February 1996 Peruvian tsunami. *Geophys. Res. Lett.* **25**, 2687–2690 (1998).
- Tichelaar, B. W. & Ruff, L. J. Depth of seismic coupling along subduction zones. *J. Geophys. Res.* **98**, 2017–2037 (1993).
- Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. Data were obtained through the IRIS datacenter. This work was supported by the US NSF. We thank S. Schwartz and E. Silver for comments on early drafts.

Correspondence and requests for materials should be addressed to S.B. (e-mail: sbilek@es.ucsc.edu).

Missing lithotroph identified as new planctomycete

Marc Strous*, John A. Fuerst†, Evelien H. M. Kramer*, Susanne Logemann*, Gerard Muyzer‡, Katinka T. van de Pas-Schoonen*, Richard Webb†, J. Gijs Kuenen* & Mike S. M. Jetten*

* Department of Biotechnology, Delft University of Technology, Julianalaan 67, 2628 BC Delft, The Netherlands

† Department of Microbiology & Parasitology and the Centre for Microscopy and Microanalysis, University of Queensland, Brisbane, Queensland 4072, Australia

‡ Netherlands Institute for Sea Research, 1790 AB Den Burg, The Netherlands

With the increased use of chemical fertilizers in agriculture, many densely populated countries face environmental problems associated with high ammonia emissions. The process of anaerobic ammonia oxidation ('anammox') is one of the most innovative technological advances in the removal of ammonia nitrogen from

waste water^{1,2}. This new process combines ammonia and nitrite directly into dinitrogen gas³. Until now, bacteria capable of anaerobically oxidizing ammonia had never been found and were known as "lithotrophs missing from nature"⁴. Here we report the discovery of this missing lithotroph and its identification as a new, autotrophic member of the order *Planctomycetales*, one of the major distinct divisions of the Bacteria⁵. The new planctomycete grows extremely slowly, dividing only once every two weeks. At present, it cannot be cultivated by conventional microbiological techniques. The identification of this bacterium as the one responsible for anaerobic oxidation of ammonia makes an important contribution to the problem of unculturability.

For over a century, microbiologists have generally acknowledged Koch's postulate that to prove that a process such as the anaerobic oxidation of ammonia is mediated by a bacterium, this bacterium should be isolated in pure culture and still be able to reproduce the process. But for the past ten years, the anammox process has resisted such meaningful microbiological characterization.

Electron microscopy provided an incentive to continue pursuing the quest for the bacterium responsible for anammox, because a peculiar morphotypical microorganism dominated the multispecies biofilms that mediated the process. Thin sectioning of cryosubstituted material revealed that cells of this morphotype were characteristically compartmented in their internal ultrastructure (Fig. 1a). Such membrane-bounded cell compartments were so far found only in cultured strains of the order *Planctomycetales*^{6,7}. When cell surfaces were examined by negative staining (Fig. 1b), the anammox morphotype displayed uniformly distributed crateriform structures, circular electron-dense areas known to be characteristic of planctomycetes⁸.

It would be surprising if a planctomycete was responsible for anammox. The order *Planctomycetales*—comprising a separate major division of the Bacteria—has been represented so far only by a few organotrophs⁸. Formerly the planctomycetes were considered to be of limited environmental importance. But this view is changing rapidly as molecular microbial ecology repeatedly provides new evidence showing that these bacteria are ubiquitous and make up a substantial portion of the natural bacterial population^{9–13}. The recognition of the planctomycetes as a major bacterial division may change our notion of what bacteria are. Planctomycetes have single- or double-membrane-bounded compartments (see above and Fig. 1) separating their chromosome from the remainder of the cytoplasm^{6,7}. They lack peptidoglycan in their cell walls and are insensitive to ampicillin^{14,15}.

As conventional purification using single-cell isolation had failed, we physically purified the morphotypical microorganism from the multispecies biofilms¹⁶ by density-gradient centrifugation. The purification procedure consisted of three steps: (1) disruption of the biofilms by mild sonication; (2) separation of the single cells obtained and the remaining biofilm fragments; and (3) purification of the single cells using Percoll density-gradient centrifugation (Table 1).

We showed that the purified cells were responsible for anaerobic ammonium oxidation by incubating the obtained cell suspensions (99.6% pure) with the substrates ammonium and nitrite (5 mM each) and measuring substrate consumption over time.

Surprisingly, it seemed that both the purified and unpurified cells were only active when the cell concentration was higher than

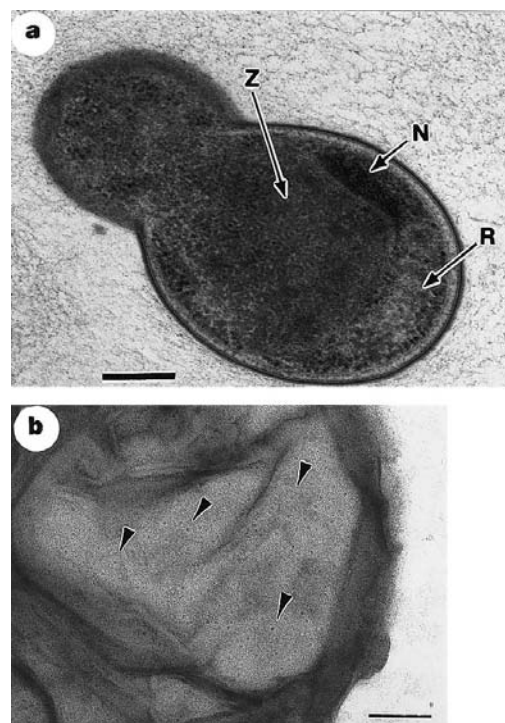


Figure 1 Transmission electron micrograph of anammox biofilms. **a**, Thin-sectioned cell of dominant morphotype from the anammox biofilm enrichment culture, demonstrating unusual compartmentalized internal organization and budding reproduction. The mother cell of the budding cell shown possesses a central membrane-bounded ribosome-free zone (Z), surrounded by a membrane-bounded ribosome-rich region (R) in which the fibrillar nucleoid (N) is situated. Bar marker is 0.2 μm . **b**, Portion of the surface of a negatively stained cell of dominant morphotype from the same culture, displaying uniformly distributed, circular crateriform structures, examples of which are indicated by arrows. Bar marker is 100 nm.

10^{10} – 10^{11} cells ml^{-1} . This was not noticed previously because we had studied anammox only in biofilms, where the cell density is naturally very high. One might argue that the high concentrations of purified cells were required to attain a sufficient amount of contaminating microorganisms required for anammox activity. However, this possibility could be eliminated because control experiments showed that the activity of unpurified cells was also density dependent (Table 1). Therefore, we concluded that the cell density of the anammox morphotype was important, and not the amount of contaminants. It is as yet unclear whether the density-dependent activity of these cells results from cell-to-cell communication through homoserine lactone autoinducers (quorum sensing)¹⁷ or whether a different mechanism is responsible for the density dependency in this case.

Because the Percoll purification procedure yielded only a small number of cells (Table 1), the anaerobic activity test was miniaturized to 30–40 μl to attain high cell densities. In addition to the required high cell concentration, the cells needed to be supplied with a trace amount of one of the anammox intermediates hydrazine or hydroxylamine (100 μM) before they became active, and the nitrite

Table 1 Purification and cell-density-dependent activity of anammox cells from anammox biofilms

Purification step	Purity* (%)	Total yield of bacterial cells (mg protein)	Specific activity* (nmol NH_4^+ per mg protein per min)	Cell density threshold (cells per ml)
Mixed culture biofilms	70 $\pm$ 10	50	25 $\pm$ 5	10^7 – 10^9 †
Single cells from sonicated biofilms	85 $\pm$ 5	2	20 $\pm$ 4	10^{10} – 10^{11}
Purified cell suspension after density gradient centrifugation	99.6 $\pm$ 0.2	0.5	18 $\pm$ 3	10^{10} – 10^{11}

*Values are expressed as mean $\pm$ standard deviation.

† Average of biofilms and culture liquid. Inside the biofilms, cells were densely packed, leading to local cell densities of at least 10^{10} – 10^{11} ml^{-1} .

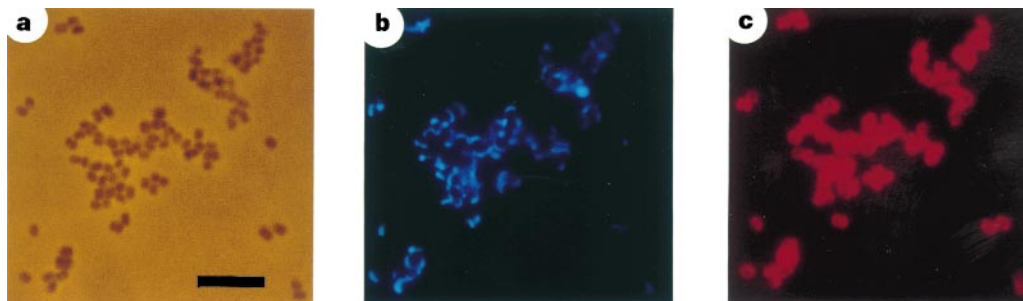


Figure 2 Fluorescent *in situ* hybridization of purified anammox cells. Identical field viewed by: **a**, phase contrast microscopy; **b**, epifluorescence microscopy with cells stained with the general DNA stain DAPI; and **c**, epifluorescence after

hybridization with anammox organism 16S rRNA-sequence based, Cy3-labelled probe S-G-Amx-1015-a-A-18 (5'-GAT ACC GTT CGT CGC CCT-3') at 20% (v/v) formamide. Bar marker is 5 μ m.

concentration needed to be lower than 5 mM to prevent substrate inhibition.

The high cell density and the low substrate concentration allowed the purified cells only a brief burst of activity, little more than 100 catabolic cycles (each cell consumed only a small amount of substrate, every individual enzyme turning over 100 times at most). Nevertheless, the activity of the purified cells was only slightly lower than the activity of the biofilms (Table 1) and the stoichiometry of the conversion was identical: 1.3 moles of nitrite were consumed and 0.2 moles of nitrate were produced per mole of ammonium consumed. Even assimilation of CO_2 by the anaerobic ammonium-oxidizing cells could be measured. During the experiment, the stoichiometry of ^{14}C - CO_2 incorporation increased rapidly from 0 to 20 millimoles of CO_2 per mole of ammonium consumed. In control experiments with unpurified cell suspensions prepared from enrichment culture biofilms, the rate of CO_2 incorporation was always less. The pathway of CO_2 assimilation is not yet resolved but the recorded ribulose biphosphate carboxylase activity in the biofilm enrichment culture was less than 50% of the activity required to account for the measured CO_2 incorporation. Alternatively, one of the other four known autotrophic pathways might be present in this case. Up until now, planctomycetes were known as organotrophic organisms (that is, incapable of autotrophic CO_2 fixation) and their biochemical pathways are so far unexplored.

Because the responsible bacterium was purified but not isolated, contaminating organisms might contribute to the observed activity. However, from kinetic and stoichiometric perspectives it is extremely unlikely that the contaminating organisms (1 in 200)

contribute significantly to anammox catabolism or anabolism. However, it is quite possible that other organisms are required for growth of the anammox organism. Possible symbionts might, for example, supply growth factors and scavenge by-products of anammox anabolism.

To determine the phylogenetic identity of the purified cells, we extracted DNA and RNA from the purified suspensions and amplified the 16S ribosomal RNA gene directly using polymerase chain reaction or indirectly after reverse transcription of RNA (RT-PCR) from complementary DNA obtained from the 16S rRNA itself. Although the cell suspensions were more than 99% pure, only the RT-PCR-derived clone libraries were dominated (80%) by a single sequence. Based on this sequence we designed 18 specific oligonucleotide probes, labelled with Cy3 fluorochrome for fluorescent *in situ* hybridization (FISH). Ten of these probes gave bright and specific hybridization signals with the purified morphotypical microorganism (Fig. 2). Combined FISH and 4,6-diamidino-2-phenylindole (DAPI) staining confirmed the internal cell compartmentalization. The FISH signal was ring shaped, indicating the absence of ribosomes in the inner cell compartment. The DAPI signal, staining the genomic DNA, was concentrated in a half-ring, which is consistent with the localization of the nucleoid in Fig. 1a.

The 16S rRNA sequence of the anammox organism (1,453 base pairs, from position 7 to 1,390; *Escherichia coli* numbering) was aligned and analysed phylogenetically with other bacterial 16S rRNA sequences¹⁸. The anammox organism was identified as a deep-branching planctomycete (Fig. 3), most closely related to a sequence amplified from a marine snow aggregate¹⁹ (sequence identity 80.2%). Sequence identity with other planctomycetes was 74–77%. The new sequence of the 16S rRNA gene shared most but not all signature positions with those of the other planctomycetes^{8,20,21}, confirming its position as a deep-branching but still related sequence.

Finally, the missing link in the biogeochemical nitrogen cycle seems to have been found; it is an organism in the order *Planctomycetales*, a distinctive division of the domain Bacteria, with unique cell organization and cell walls, large evolutionary distances and few cultivated representatives.

Methods

Purification and activity. Biofilms¹⁶ were washed and concentrated in 40 ml HEPES/bicarbonate buffer (70/10 mM, pH 7.8), sonicated in 5 ml portions (150 W, 30 s, 18-mm tip width) and, after centrifugation (10,000g, 5 min), an upper orange pellet containing undisrupted biofilm fragments was separated from the lower red pellet containing the single cells. After washing in the same buffer, the single-cell fraction was purified using Percoll centrifugation (6.9 ml Percoll (Pharmacia) and 3.1 ml cell suspension; centrifugation at 10,000g for 60 min at 5°C; Sorval SS34 fixed angle rotor). A broad, red band of target cells in the lower half of the tube was extracted and washed in sterile buffer (see above). Anammox activity was tested in a 30- μ l activity test in anaerobic containers

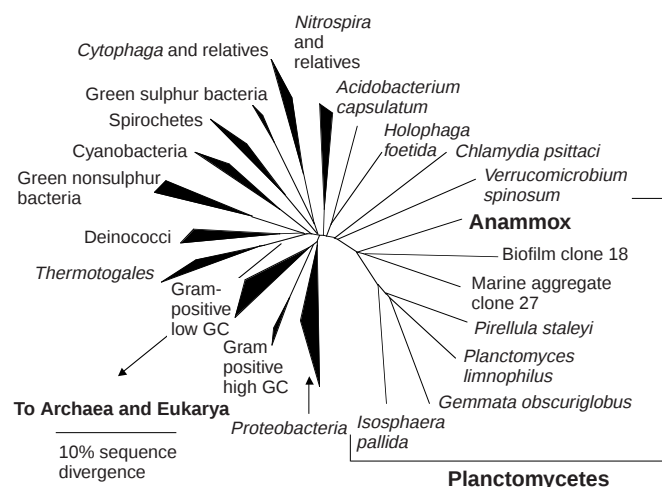


Figure 3 Phylogenetic position of the lithotroph responsible for anaerobic ammonium oxidation within the domain Bacteria, based on 16S rRNA phylogeny. The anammox bacterium represents a new, deep branch inside the order *Planctomycetales*.

(same buffer containing 4.5 mg protein ml⁻¹ (10¹⁰–10¹¹ cells ml⁻¹), 5 mM ammonium and nitrite, 100 µM hydrazine) and consumption of ammonium and nitrite and production of nitrate and incorporation of ¹⁴C-CO₂ was measured as described previously^{16,22}. Purity was assessed by microscopic counting, electron microscopy and FISH.

Electron microscopy. Pelleted cell flocs were processed by a cryosubstitution protocol before embedding, sectioning and section staining, all using methods described in ref. 7. Cells were negatively stained using 1% uranyl acetate + 0.4% sucrose, after dispersal of flocs.

Fluorescent microscopy. FISH and DAPI staining were done as described¹⁰. Formamide concentration versus specificity of all probes was determined with the appropriate reference organism for each probe and with the biofilm enrichment culture and several undefined, heterogeneous environmental samples. All probes were at least 18 nucleotides long.

Phylogeny. Treeing and phylogenetic analysis was done using the ARB software package¹⁸. The sequence was obtained after DNA isolation, PCR amplification (primers 27 or 519 forward and 1,390 reverse; annealing at 46 °C, melting at 94 °C, 30 cycles), ligation of the product in vector pGEMT (Promega) and transformation into competent DH5α *E. coli* cells. The clone library was screened using restriction digestion, and representative clones of each restriction pattern were sequenced. The anammox sequence (GenBank accession number AJ131819) always grouped with the other planctomycete sequences, independent of treeing algorithm (neighbour joining, distance matrix or parsimony), inclusion of other bacterial phyla in the tree or the choice of the outgroup (*Thermotoga*, *Aquifex* or the Archaea).

Received 4 May; accepted 1 June 1999.

1. Jetten, M. S. M., Horn, S. J. & van Loosdrecht, M. C. M. Towards a more sustainable wastewater treatment system. *Wat. Sci. Tech.* **35**, 171–180 (1997).
2. Strous, M., Van Gerven, E., Ping, Z., Kuenen, J. G. & Jetten, M. S. M. Ammonium removal from concentrated waste streams with the anaerobic ammonium oxidation process. *Water Res.* **31**, 1955–1962 (1997).
3. Van de Graaf, A. A., de Bruijn, P., Robertson, L. A., Jetten, M. S. M. & Kuenen, J. G. Metabolic pathway of anaerobic ammonium oxidation on basis of ¹⁵N-studies in a fluidized bed reactor. *Microbiology* **143**, 2415–2421 (1997).
4. Broda, E. Two kinds of lithotrophs missing in nature. *Z. Allgem. Mikrobiol.* **17**, 491–493 (1977).
5. Schlesner, H. & Stackebrandt, E. Assignment of the genera *Planctomyces* and *Pirella* to a new family *Planctomycetaceae* fam. nov. and description of the order *Planctomycetales* ord. nov. *Syst. Appl. Microbiol.* **8**, 174–176 (1986).
6. Fuerst, J. A. & Webb, R. I. Membrane-bounded nucleoid in the eubacterium *Gemmata obscuriglobus*. *Proc. Natl Acad. Sci. USA* **88**, 8184–8188 (1991).
7. Lindsay, M. R., Webb, R. I. & Fuerst, J. A. Pirellulosomes: a new type of membrane-bounded cell compartment in planctomycete bacteria of the genus *Pirella*. *Microbiology* **143**, 739–748 (1997).
8. Fuerst, J. A. The planctomycetes: emerging models for microbial ecology, evolution and cell biology. *Microbiology* **141**, 1493–1506 (1995).
9. Vergin, K. L. et al. Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order *Planctomycetales*. *Appl. Environ. Microbiol.* **64**, 3075–3078 (1998).
10. Neef, A., Amann, R. L., Schlesner, H. & Schleifer, K. H. Monitoring a widespread bacterial group: in situ detection of planctomycetes with 16S rRNA-targeted probes. *Microbiology* **144**, 3257–3266 (1998).
11. Hugenholtz, P., Goebel, B. M. & Pace, N. R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**, 4765–4774 (1998).
12. Byers, H. K., Stackebrandt, E., Hayward, C. & Blackall, L. L. Molecular investigation of a microbial mat associated with the Great Artesian Basin. *FEMS Microbiol. Ecol.* **25**, 391–403 (1998).
13. Zarda, B. et al. Analysis of bacterial community structure in bulk soil by in situ hybridization. *Arch. Microbiol.* **168**, 185–192 (1997).
14. Liesack, W., König, H., Schlesner, H. & Hirsch, P. Chemical composition of the peptidoglycan-free cell envelopes of budding bacteria of the *Pirella/Planctomyces* group. *Arch. Microbiol.* **145**, 361–366 (1986).
15. König, E., Schlesner, H. & Hirsch, P. Cell wall studies on budding bacteria of the *Planctomyces/Pasteuria* group and on a *Prosthecomicrobium* sp. *Arch. Microbiol.* **138**, 200–205 (1984).
16. Strous, M., Heijnen, J. J., Kuenen, J. G. & Jetten, M. S. M. The sequencing batch reactor as a powerful tool for the study of slowly growing anaerobic ammonium oxidizing microorganisms. *Appl. Microbiol. Biotechnol.* **50**, 589–596 (1998).
17. Fuqua, C. & Greenberg, E. Self perception in bacteria: quorum sensing with acylated homoserine lactones. *Curr. Opin. Microbiol.* **1**, 183–189 (1998).
18. Ludwig, W. et al. Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis* **19**, 554–568 (1998).
19. Delong, E. F., Franks, D. G. & Alldredge, A. L. Phylogenetic diversity of aggregate-attached versus free-living marine bacterial assemblages. *Limnol. Oceanogr.* **38**, 924–934 (1993).
20. Fuerst, J. A. et al. Isolation and molecular identification of planctomycete bacteria from postlarvae of the giant tiger prawn *Penaeus monodon*. *Appl. Environ. Microbiol.* **63**, 254–262 (1997).
21. Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51**, 221–271 (1987).
22. Van de Graaf, A. A., Debruijn, P., Robertson, L. A., Jetten, M. S. M. & Kuenen, J. G. Autotrophic growth of anaerobic ammonium oxidizing microorganisms in a fluidized bed reactor. *Microbiology* **142**, 2187–2196 (1996).

Acknowledgements. We thank A. Raghoebarsing and J. Schalk for technical assistance; M. Wagner (TU Munich, Germany) for instruction on fluorescent *in situ* hybridization; and the Foundation of Applied Research (STW) for financial support. J.A.F. and R.W. acknowledge the support of the Australian Research Council.

Correspondence and requests for materials should be addressed to M.S.M.J. (e-mail: m.jetten@stm.tudelft.nl). The anammox sequence has been deposited in GenBank (accession no. AJ131819).

Sperm competition between *Drosophila* males involves both displacement and incapacitation

Catherine S. C. Price, Kelly A. Dyer* & Jerry A. Coyne

Department of Ecology and Evolution, The University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA

Females in almost all animal groups copulate with multiple males^{1,2}. This behaviour allows different males to compete for fertilization³ and gives females the opportunity to mediate this competition⁴. In many animals and most insects, the second male to copulate with a female typically sires most of her offspring^{1,5,6}. In *Drosophila melanogaster*, this second-male sperm precedence has long been studied^{7–15} but, as in most species, its mechanism has remained unknown. Here we show, using labelled sperm in doubly mated females, that males can both physically displace and incapacitate stored sperm from earlier-mating males. Displacement occurs only if the second male transfers sperm to the female, and in only one of her three sperm-storage organs. Incapacitation can be caused by either fertile or spermless second males, but requires extended intervals between matings. Sperm from different males are not ‘stratified’ in the storage organs but mix freely. Many animal species may have multiple mechanisms of sperm competition like those observed here, and revealing these mechanisms is necessary to understand the genetic and evolutionary basis of second-male sperm precedence in animals.

The consequences of multiple mating are well known in *D. melanogaster*. Females store sperm in a long tubular seminal receptacle and two mushroom-shaped spermathecae¹⁰. The proportion of offspring produced after the second mating that are sired by the second male, *P*₂, is typically above 0.8 in laboratory experiments¹⁶. Most wild-caught females carry sperm from multiple males¹⁷, and females in nature may copulate when they still carry a sizeable sperm load from previous matings¹⁸. If a sperm-carrying female remates, she produces fewer offspring sired by the first male than she would have if she had not remated¹⁶. At least some of this reduction in first-male reproductive success occurs if the second male transfers only seminal fluid and no sperm^{12,14}. Seminal fluid therefore plays some role in second-male sperm precedence¹⁵, but the extent of this role is unclear because some studies^{10,19} have not revealed a ‘seminal-fluid effect’. To investigate this discrepancy, we looked for a seminal-fluid effect in experiments carried out at two different time intervals.

Rematings to fertile second males produced the usual result of second-male sperm precedence and resulted in significant reduction of first-male reproductive success when females remated after either two or seven days (Fig. 1a). When females were remated in exactly the same manner to infertile XO males transferring only seminal fluid (see Methods), there was a significant reduction in first-male offspring when females were remated after seven days, but only a slight and non-significant reduction when they were remated after two days (Fig. 1a). Similar results were found when this experiment was repeated with Ives males instead of *bw*^D males (carrying the brown-dominant mutation) as first mates (data not shown). These results confirm that at least some of the second male’s precedence can be ascribed to his seminal fluid, and, because this effect was not seen until seven days after remating, indicate that first-male sperm

* Present address: Department of Biology, University of Rochester, Rochester, New York 14627, USA.

must remain in storage for a few days before they become susceptible to the harmful effects of second-male seminal fluid. Second-male precedence at shorter remating intervals apparently requires the transfer of sperm. This indicates at least two mechanisms of sperm precedence in *D. melanogaster*, only one of which requires sperm in the ejaculate.

The loss of first-male productivity caused by second-male seminal fluid was not accompanied by any loss of first-male sperm from storage (Fig. 1b). Although females that copulate with XO males seven days after their first mating produce about one-third as many offspring as they would have if they had not remated (Fig. 1a), XO-remated females have no fewer sperm in storage than do singly mated females (Fig. 1b). This indicates that the seminal fluid of a second male inhibits the use of stored sperm without removing them. Moreover, we found no evidence that the paucity of first-male progeny can be attributed to developmental problems in eggs fertilized by first-male sperm. Mean ($\pm$ s.e.) egg hatchability for the first 48 h after mating was 0.85 ($\pm$ 0.05; $n = 14$) for females singly mated to an Ives male; 0.84 ($\pm$ 0.03; $n = 28$) for females singly mated to a *bw^D* male; and 0.87 ($\pm$ 0.03; $n = 18$) for females mated first to a *bw^D* and then to an Ives male. Thus, interference with the effectiveness of first-male sperm occurs before and not after fertilization. We refer to this interference as 'sperm incapacitation'⁶.

Each of the experiments reported above was also performed in *Drosophila simulans*, a sister species of *D. melanogaster*, with similar results (data not shown). This indicates that incapacitation of stored sperm after extended remating intervals may be widespread in *Drosophila*.

To study the mechanics of sperm precedence when second males are fully fertile, we mated females to males that express green fluorescent protein (GFP) on their sperm tails²⁰ and then to males that lack GFP expression but carry a dominant eye-colour mutation, brown-dominant (*bw^D*). In this way, each offspring produced and each sperm stored by such a female could be assigned to one of her two mates (Fig. 2). Preliminary experiments confirmed that GFP males transfer only labelled sperm to females. In six females singly mated to GFP males, we found 516 GFP-labelled sperm and no unlabelled sperm. When GFP males copulate first, sperm precedence is high and within the normal range for *D. melanogaster* (Fig. 3a). Furthermore, GFP males suffer a net reproductive loss when females remate to *bw^D* males after either four or seven days (Fig. 3a). Most of the stored sperm in these females comes from the second male (Fig. 3b). Regardless of the interval between matings, doubly mated females have fewer first-male sperm in the seminal receptacle than do single-mated females (Fig. 3c). In contrast, first-

male sperm stored in the paired spermathecae show a much smaller and non-significant diminution after double matings (Fig. 3d).

It has been suggested that females lack the space to store two ejaculates without expelling first-male sperm¹⁰. There is, in fact, more than enough space for two ejaculates, and we observed loss of first-male sperm from the seminal receptacle despite this excess capacity. Females dissected up to 48 h after a single mating with a *bw^D* male had a mean of 256 ($\pm$ 26.9; $n = 15$) and as many as 377 sperm in their seminal receptacles. On the fourth day after mating, those females had a mean of 63.4 ($\pm$ 16.3; $n = 5$) sperm, and on the seventh day only 24.3 ($\pm$ 8.2; $n = 5$) sperm in their seminal receptacles. If a female can store over 250 sperm, but carries an average of only 24 sperm at the time of remating, sperm displacement is not required to make space for the second ejaculate, nor even for second-male sperm to make up the majority in the seminal receptacle (in such a case with no displacement, P_2 would be $250/(24 + 250) = 0.91$).

Each of the experiments reported above was repeated with the mating order of GFP and *bw^D* males reversed. The GFP males are not as good at sperm offence as the *bw^D* males, and little sperm precedence was found with four days between matings ($P_2 = 0.54 \pm 0.10$; $n = 12$). However, normal sperm precedence was found with seven days between matings ($P_2 = 0.77 \pm 0.04$; $n = 15$). Comparing the number of first-male sperm stored in females remated after seven days with the number stored in singly mated females at the same time, we find that GFP second males caused physical loss of *bw^D* sperm from the seminal receptacle (one-tailed Mann-Whitney *U*, $P = 0.002$) but not from the spermathecae (one-tailed Mann-Whitney *U*, $P = 0.378$). The magnitude of the loss of sperm from the seminal receptacle was compared to the loss from the spermathecae using a two-tailed *t*-test²¹ for each of the four double matings ($P = 0.050$ for GFP then *bw^D* 4-day interval; $P = 0.264$ for GFP then *bw^D* 7-day interval; $P = 0.059$ for *bw^D* then GFP 4-day interval; $P = 0.036$ for *bw^D* then GFP 7-day interval). Combining the probabilities from all four independent comparisons²², the loss of sperm from the seminal receptacle was significantly greater than the loss of sperm from the spermathecae ($P = 0.0072$).

None of the dissections of doubly mated females supports the hypothesis²³ that sperm are layered in storage, so that second-male sperm are closer to the site of fertilization. Sperm masses were removed intact from the storage organs, and GFP and non-GFP sperm were always found closely intertwined.

Most proposed mechanisms of second-male sperm precedence in *D. melanogaster* have sought to explain how second-male sperm

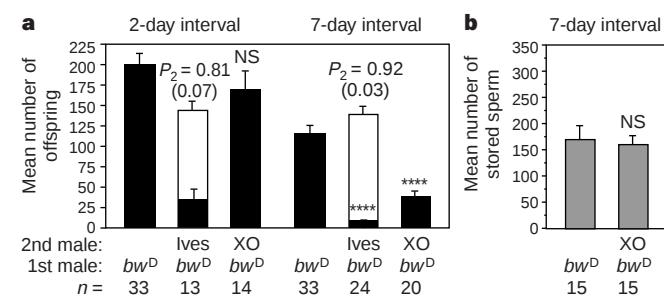


Figure 1 The mechanism of sperm competition depends on mating interval. **a**, Mean ($\pm$ s.e.) number of offspring per female from first (solid bars) and second (open bars) matings. Offspring from single matings are those produced after the doubly mated females remated. P_2 (s.e.) is the proportion of second-male offspring produced after remating. **** $P < 0.0001$; NS: $P = 0.113$ (one-tailed Mann-Whitney *U* test, compared to the singly mated control). **b**, Mean ($\pm$ s.e.) number of sperm stored per female in all organs. Data are pooled from females dissected 7, 8 and 9 days after the first mating. $P = 0.459$ (one-tailed Mann-Whitney *U*-test).

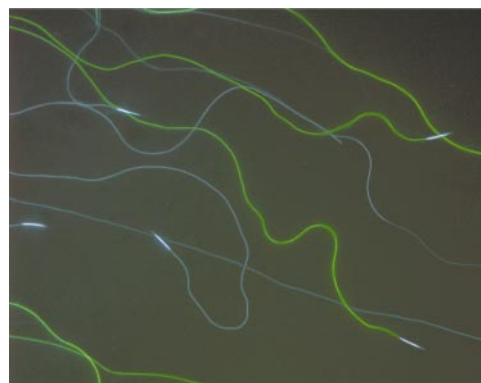


Figure 2 Differentially labelled sperm. DAPI-stained sperm from a GFP male (green tails) and a *bw^D* male (blue tails), dissected from the seminal receptacle of a single doubly inseminated female.

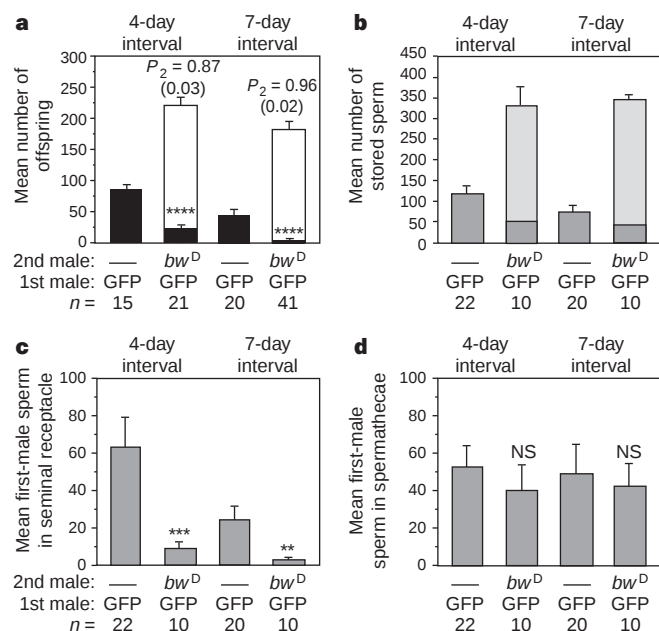


Figure 3 Physical displacement of first-male sperm. **a**, Mean (+ s.e.) number of offspring per female from first (solid bars) and second (open bars) matings. **** $P < 0.0001$. **b**, Mean (+ s.e.) number of sperm stored per female in all organs from first (dark bars) and second (light bars) matings. **c**, Mean (+ s.e.) number of sperm stored per female in the seminal receptacle. *** $P = 0.001$; ** $P = 0.006$. **d**, Mean (+ s.e.) number of sperm stored per female in the spermathecae. $P = 0.412$ for the 4-day interval; $P = 0.387$ for the 7-day interval (one-tailed Mann-Whitney U -tests).

achieve numerical superiority in storage^{9–11,14}. It is often supposed that patterns of paternity directly reflect such a numerical advantage, implying that, once sperm storage is complete, females use the available sperm at random to fertilize eggs. We confirm here that the majority of stored sperm after double matings are second-male sperm (Fig. 3b), and attribute this in part to physical displacement of first-male sperm (Fig. 3c). These observations do not, however, rule out the possibility that females preferentially use first- or second-male stored sperm to fertilize their eggs.

Table 1 compares the number of sperm a female stores from a given male with the number of offspring ultimately produced by that male, and estimates the proportion of each type of stored sperm that is actually used to fertilize eggs. Seven days after a single mating to a bw^D male, females use about 68% of their remaining stored sperm to produce offspring (Table 1). If that female is remated to an XO male seven days after mating to the bw^D male, she then uses only about 24% of her stored sperm to produce offspring. This reflects

the sperm incapacitation described above. Females who copulate first with a GFP male and then after four days with a bw^D male subsequently use about 48% of their first-male sperm to produce offspring (Table 1). However, with seven days between matings, females use only about 7% of the stored first-male sperm. While the number of first-male sperm stored by females remated after seven days does not differ significantly from the number stored by females remated after four days, the former produce significantly fewer offspring from the first male (Table 1). Sperm precedence is thus very high after the seven-day remating ($P_2 = 0.96 \pm 0.02$; $n = 41$), almost certainly because many first-male sperm are displaced from the seminal receptacle (Fig. 3c), and because most of the sperm that survive displacement are incapacitated. Again, the incapacitation effect is seen only with the longer interval between matings.

Sperm competition offers a unique opportunity to study adaptations shaped by the interacting forces of natural, sexual and sexually antagonistic selection²⁴. Future work in these species must explain, both mechanistically and evolutionarily, why first-male sperm become more susceptible to incapacitation the longer they remain in storage. This phenomenon may explain how males avoid incapacitating the sperm in their own ejaculate, despite their ability to displace and/or incapacitate their own sperm once it has been stored in a female¹⁵. Females may play a role in sperm incapacitation if they gradually alter the physiology of the sperm they store. Genetic differences among females in the ability to alter sperm in this way might explain why female genotypes differ widely in the degree of last-male precedence²⁵. Moreover, if females evolve differences in the way they alter stored sperm and males evolve ways to protect their sperm from alteration by particular females, this might explain the observation that a first male's success depends on both his genotype and that of his mate²⁶. Sperm competition in animals other than *Drosophila* may involve mechanisms similar to those described here, and any general explanation of the ubiquity of second-male sperm precedence among animals must consider that multiple mechanisms may have evolved within a single species. □

Methods

Matings. Stock maintenance, mating observations and the rearing of offspring were performed as described²⁷. All females came from the Ives laboratory population²⁸ and males were taken as noted from the Ives, brown-dominant (bw^D) or *djGFP II/CyO* (GFP) stock²⁰. XO males were produced by crossing virgin Ives females to males from an attached-X, attached-XY stock. All male offspring from this cross lack a Y chromosome, and produce normal seminal fluids but no sperm^{29,30}. Females were transferred to fresh food vials every three days until they either stopped laying fertile eggs or were chosen at random for dissection. All offspring were reared to adulthood and scored for paternity, determined by the presence of brown or wild-type eyes.

Dissections. For each mating type, 5–10 females were dissected per day at timed intervals after the end of their last copulation. Females were etherized and their reproductive tracts removed in a drop of phosphate buffered saline (PBS). The spermathecae, seminal receptacle and uterus were each transferred

Table 1 Proportion of sperm used to produce offspring

1st male	2nd male	Days	First male			Second male		
			Offspring produced	Sperm stored	Sperm used (%)	Offspring produced	Sperm stored	Sperm used (%)
bw^D	—	—	114.1 (12.1)	168.8 (26.9)	68%	—	—	—
			****	NS				
bw^D	XO	7	37.9 (7.3)	157.8 (22.3)	24%	—	—	—
GFP	bw^D	4	23.7 (4.2)	49.8 (35.0)	48%	197.5 (13.6)	300.2 (87.6)	66%
			****	NS				
GFP	bw^D	7	3.0 (0.7)	43.2 (14.3)	7%	183.2 (13.1)	338.4 (19.5)	54%

Days are the number of days between the first and second mating. Per cent sperm used is calculated by dividing the mean number of offspring produced by the mean number of sperm stored. Mean (s.e.) of offspring and sperm are from data shown in Figs 1 and 3. Data for the single mating are from beyond the first seven days after mating. One-tailed Mann-Whitney U -tests: **** $P < 0.0001$; NS (non-significant) $P > 0.4$.

to a separate drop of PBS to prevent cross-contamination of sperm from different organs. The sperm from each organ were then removed with insect pins and minimal manipulation. Slides to be scored for GFP vs non-GFP sperm were dissected in $0.5 \mu\text{g ml}^{-1}$ DAPI (4,6-diamidino-2-phenylindole) in PBS, covered immediately with a coverslip and scored with an epifluorescent microscope within 2 h of dissection. Slides that required only simple sperm counts were dried after dissection at 60°C for 5–10 min, fixed in 3:1 methanol and glacial acetic acid for 5 min, rinsed three times in PBS and labelled with DAPI in glycerol ($0.5 \mu\text{g ml}^{-1}$). All sperm in each of the three storage organs of every female were counted. To ensure precision, 10% of the slides were counted at least twice by different people.

Egg hatchability. Females were allowed to lay eggs on small plastic spoons filled with grape-juice-tinted medium for 24 h, and then either transferred to a fresh spoon or discarded. Spoons were stored at 24°C for 28 h before hatched and unhatched eggs were counted. Brown eggs, which may indicate zygotes that died early in development, were observed only rarely.

Received 3 March; accepted 28 May 1999.

- Smith, R. L. (ed.) *Sperm Competition and the Evolution of Animal Mating Systems* (Academic, London, 1984).
- Birkhead, T. R. & Møller, A. P. (eds.) *Sperm Competition and Sexual Selection* (Academic, London, 1998).
- Parker, G. A. Sperm competition and its evolutionary consequences in the insects. *Biol. Rev.* **45**, 525–567 (1970).
- Eberhard, W. G. *Female Control: Sexual Selection by Cryptic Female Choice* (Princeton Univ. Press, Princeton, 1996).
- Birkhead, T. R. & Møller, A. P. *Sperm Competition in Birds: Evolutionary Causes and Consequences* (Academic, London, 1992).
- Simmons, L. W. & Siva-Jothy, M. T. in *Sperm Competition and Sexual Selection* (eds Birkhead, T. R. & Møller, A. P.) 341–434 (Academic, London, 1998).
- Nonidez, J. F. The internal phenomena of reproduction in *Drosophila*. *Biol. Bull.* **39**, 207–230 (1920).
- Lobashov, M. E. Mixture of sperm in case of polyandry in *Drosophila melanogaster*. *C. R. (Doklady) Acad. Sci. de l'URSS* **23**, 827–830 (1939).
- Kaufmann, B. P. & Demerec, M. Utilization of sperm by the female *Drosophila melanogaster*. *Am. Nat.* **76**, 445–469 (1942).
- Lefevre, G. & Jonsson, U. B. Sperm transfer, storage, displacement, and utilization in *Drosophila melanogaster*. *Genetics* **47**, 1719–1736 (1962).
- Newport, M. E. A. & Gromko, M. H. The effect of experimental design on female receptivity to remating and its impact on reproductive success in *Drosophila melanogaster*. *Evolution* **38**, 1261–1272 (1984).
- Scott, D. & Richmond, R. C. Sperm loss by remating *Drosophila melanogaster* females. *J. Insect Physiol.* **36**, 451–456 (1990).
- Scott, D. & Williams, E. Sperm displacement after remating in *Drosophila melanogaster*. *J. Insect Physiol.* **39**, 201–206 (1993).
- Harshman, L. G. & Prout, T. Sperm displacement without sperm transfer in *Drosophila melanogaster*. *Evolution* **48**, 758–766 (1994).
- Gilchrist, A. S. & Partridge, L. Male identity and sperm displacement in *Drosophila melanogaster*. *J. Insect Physiol.* **41**, 1087–1092 (1995).
- Gromko, M. H., Gilbert, D. G. & Richmond, R. C. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 371–426 (Academic, London, 1984).
- Imhof, M., Harr, B., Brem, G. & Schlötterer, C. Multiple mating in wild *Drosophila melanogaster* revisited by microsatellite analysis. *Mol. Ecol.* **7**, 915–917 (1998).
- Gromko, M. H. & Markow, T. A. Courtship and remating in field populations of *Drosophila*. *Anim. Behav.* **45**, 253–262 (1993).
- Gromko, M. H., Newport, M. E. A. & Kortier, M. G. Sperm dependence of female receptivity to remating in *Drosophila melanogaster*. *Evolution* **38**, 1273–1282 (1984).
- Santel, A., Winbauer, T., Blümler, N. & Renkawitz-Pohl, R. The *Drosophila don juan (dj)* gene encodes a novel sperm specific protein component characterized by an unusual domain of a repetitive amino acid motif. *Mech. Dev.* **64**, 19–30 (1997).
- Snedecor, G. W. & Cochran, W. G. *Statistical Methods* (Iowa State Univ. Press, Ames, 1967).
- Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, New York, 1995).
- Nachtsheim, H. Eine Methode zur Prüfung der Lebensdauer genotypischer verschiedener Spermen bei *Drosophila*. Verhandl. V. Intern. Kongr. Vererbungsw., 1143–1147. *Z. Ind. Abst. Vererb.* (Suppl. II) (1927).
- Rice, W. R. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature* **381**, 232–234 (1996).
- Clark, A. G. & Begun, D. J. Female genotypes affect sperm displacement in *Drosophila*. *Genetics* **149**, 1487–1493 (1998).
- Clark, A. G., Begun, D. J. & Prout, T. Female × male interactions in *Drosophila* sperm competition. *Science* **283**, 217–220 (1999).
- Price, C. S. C. Conspecific sperm precedence in *Drosophila*. *Nature* **388**, 663–666 (1997).
- Coyne, J. A., Aulard, S. & Berry, A. Lack of underdominance in a naturally occurring pericentric inversion in *Drosophila melanogaster* and its implications for chromosome evolution. *Genetics* **129**, 791–802 (1991).
- Ingman-Baker, J. & Candido, E. P. Proteins of the *Drosophila melanogaster* male reproductive system: two-dimensional gel patterns of proteins synthesized in the XO, XY, and YXY testis and paragonial gland and evidence that the Y chromosome does not code for structural sperm proteins. *Biochem. Genet.* **18**, 809–828 (1980).
- Kiefer, B. I. Ultrastructural abnormalities in developing sperm of X/O *Drosophila melanogaster*. *Genetics* **54**, 14441–1452 (1966).

Acknowledgements. We thank M. DeAngelis, C. Gronlund, C. Kim, C. Mercader, J. Postuszny and A. Travelli for technical assistance; A. Santel for the *dfjGFP II/Cyo* flies; A. James and R. Snook for miscellaneous help; B. Robbins for supplies; and A. Civetta, T. Prout and C.-I. Wu for comments. This work was supported by an NSF predoctoral fellowship, NIH genetics training grant and NSF doctoral dissertation improvement grant to C.S.C.P. and by an NIH grant to J.A.C.

Correspondence and requests for materials should be addressed to J.A.C. (e-mail: j-coyne@uchicago.edu).

Vanilloid receptors on sensory nerves mediate the vasodilator action of anandamide

Peter M. Zygmunt^{*†}, Jesper Petersson[†], David A. Andersson[†], Huai-hu Chuang[‡], Morten Sørsgård[†], Vincenzo Di Marzo[§], David Julius[‡] & Edward D. Högestätt^{*†}

[†] Department of Clinical Pharmacology, Institute of Laboratory Medicine, University of Lund, S-221 85 Lund, Sweden

[‡] Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94143-0450, USA

[§] C.N.R. Istituto per la Chimica di Molecole di Interesse Biologico, Via Toiano 6, Arco Felice, Napoli 80072, Italy

* These authors contributed equally to this work

The endogenous cannabinoid receptor agonist anandamide¹ is a powerful vasodilator of isolated vascular preparations^{2–4}, but its mechanism of action is unclear. Here we show that the vasodilator response to anandamide in isolated arteries is capsaicin-sensitive and accompanied by release of calcitonin-gene-related peptide (CGRP). The selective CGRP-receptor antagonist 8-37 CGRP (ref. 5), but not the cannabinoid CB1 receptor blocker SR141716A (ref. 7), inhibited the vasodilator effect of anandamide. Other endogenous (2-arachidonylglycerol, palmitylethanolamide) and synthetic (HU 210, WIN 55,212-2, CP 55,940) CB1 and CB2 receptor agonists¹ could not mimic the action of anandamide. The selective 'vanilloid receptor' antagonist capsazepine^{6,7} inhibited anandamide-induced vasodilation and release of CGRP. In patch-clamp experiments on cells expressing the cloned vanilloid receptor (VR1)⁸, anandamide induced a capsazepine-sensitive current in whole cells and isolated membrane patches. Our results indicate that anandamide induces vasodilation by activating vanilloid receptors on perivascular sensory nerves and causing release of CGRP. The vanilloid receptor may thus be another molecular target for endogenous anandamide, besides cannabinoid receptors, in the nervous and cardiovascular systems.

Anandamide (arachidonylethanolamide) was originally isolated from brain as an endogenous cannabinoid receptor ligand⁹. Bio-synthetic pathways for anandamide are also present outside the central nervous system, for example, in vascular endothelium and macrophages^{10,11}. It has been suggested that anandamide induces hypotension in anaesthetized rats by inhibiting peripheral sympathetic neurotransmission¹². Macrophage-derived anandamide is implicated in haemorrhagic shock¹³ and endotoxin-induced hypotension¹⁴. Although CB1 receptor messenger RNA has been detected in sympathetic nerves, vascular endothelium and smooth muscle^{10,15,16}, the involvement of cannabinoid receptors in the vasodilator effects of anandamide in isolated vascular preparations has been questioned^{12,4,17}. Anandamide is structurally related to capsaicin and olvanil (*N*-vanillyloleamide), compounds that all have an amide bond and an aliphatic side chain. Capsaicin and olvanil activate a subpopulation of primary sensory neurons, which can then become refractory to subsequent stimuli (desensitization)⁷. As such nerves mediate vasodilation¹⁸, we hypothesized that the vascular effects of anandamide and capsaicin have a common mechanism, involving excitation of primary sensory nerves in the vessel wall and consequent release of vasodilator neuropeptides such as CGRP.

To test this proposal, we examined the effects of capsaicin and the selective CGRP-receptor antagonist 8-37 CGRP (ref. 5) on anandamide-induced relaxation in rat hepatic and small mesenteric arteries and guinea-pig basilar artery. Pretreatment with capsaicin to cause desensitization and/or neurotransmitter depletion in perivascular sensory nerves abolished anandamide-induced relaxation in all

three arteries (Fig. 1a, b). 8-37 CGRP also inhibited the response to anandamide (Fig. 1c). As shown in the rat hepatic artery, the relaxation induced by exogenous CGRP was unaffected by pretreatment with capsaicin (10 μ M), whereas 8-37 CGRP (2 μ M) caused a significant rightward parallel shift of the concentration–relaxation curve for CGRP (control: $pEC_{50} = 9.56 \pm 0.04$, $E_{max} = 100 \pm 1\%$; capsaicin: $pEC_{50} = 9.51 \pm 0.02$, $E_{max} = 100 \pm 1\%$; 8-37 CGRP: $pEC_{50} = 8.54 \pm 0.13$, $E_{max} = 100 \pm 1\%$; $n = 6$).

At a concentration that induced vasodilation, anandamide elicited significant release of CGRP and an increase in tissue content of cyclic AMP in the rat hepatic artery (Fig. 2a, b), the latter effect presumably reflecting activation of CGRP receptors on the vasculature. Both these effects were absent in arterial segments pretreated with capsaicin (Fig. 2a, b). However, pretreatment with capsaicin did not affect the cyclic AMP increase evoked by exogenous CGRP (Fig. 2b), which is consistent with capsaicin acting on sensory nerves. CGRP-like immunoreactive nerve fibres were visualized in the rat hepatic artery (Fig. 2c). Such nerves are also present in rat mesenteric¹⁸ and guinea-pig basilar¹⁹ arteries. Together, these findings support our proposal that CGRP released from perivascular sensory nerves mediates vasodilation by anandamide. The pronounced vasodilator effect of anandamide indicates that these nerves represent a powerful effector system for regulation of vascular tone.

To investigate the role of cannabinoid receptors in these vascular responses, we compared the effects of anandamide with those of a series of endogenous (palmitylethanolamide, 2-arachidonylglycerol) and synthetic (HU 210, WIN 55,212-2, CP 55,940) cannabinoid receptor agonists with varying selectivity for CB1 and CB2 receptors¹. HU 210, WIN 55,212-2 and CP 55,940 are at least one order of magnitude more potent than anandamide as agonists of CB1 and CB2 receptors¹. In rat hepatic and guinea-pig basilar arteries, palmitylethanolamide, HU 210 and WIN 55,212-2 did

not produce any relaxation, whereas 2-arachidonylglycerol induced a small (rat hepatic artery) or inconsistent (guinea-pig basilar artery) relaxation at 10 μ M (Fig. 3). CP 55,940 elicited significant relaxation in both arteries, but this effect was resistant to pretreatment with capsaicin and occurred only at a concentration of 10 μ M (Fig. 3c, d). The failure of these different cannabinoid receptor agonists to mimic the action of anandamide indicates that CB1 or CB2 receptors do not mediate anandamide-induced vasorelaxation in these arteries.

The competitive CB1 receptor antagonist SR141716A (0.3 μ M), which binds to this receptor with a K_i in the nanomolar range¹, did not attenuate the vasodilator responses to anandamide in rat hepatic (control: $pEC_{50} = 6.45 \pm 0.11$, $E_{max} = 100 \pm 1\%$; SR141716A: $pEC_{50} = 6.44 \pm 0.12$, $E_{max} = 97 \pm 1\%$; $n = 5$) and guinea-pig basilar (control: $pEC_{50} = 6.02 \pm 0.11$, $E_{max} = 92 \pm 3\%$; SR141716A: $pEC_{50} = 6.11 \pm 0.10$, $E_{max} = 91 \pm 2\%$; $n = 5-8$) arteries, and to methanandamide in the rat hepatic artery (control: $pEC_{50} = 6.50 \pm 0.08$, $E_{max} = 99 \pm 1\%$; SR141716A: $pEC_{50} = 6.30 \pm 0.04$, $E_{max} = 99 \pm 1\%$; $n = 5$). High concentrations of SR141716A ($\geq 3 \mu$ M) can inhibit anandamide-induced relaxation in rat hepatic and mesenteric arteries, but the mechanism of this action is unclear^{3,17}. Such high concentrations of SR141716A have effects unrelated to inhibition of CB1 or CB2 receptors^{20,21}. This is also supported by this study, showing that the vasodilator response to capsaicin, which does not bind to CB1 receptors²², is significantly inhibited by 10 μ M SR141716A in the rat hepatic artery (control: $pEC_{50} = 8.37 \pm 0.13$, $E_{max} = 99 \pm 1\%$; SR141716A: $pEC_{50} = 7.42 \pm 0.12$, $E_{max} = 90 \pm 3\%$; $n = 4$).

It has been suggested that enzymatic cleavage of anandamide by a specific amidohydrolase to arachidonic acid and ethanolamine²², with subsequent formation of vasodilator eicosanoids such as prostacyclin and eicosatrienoic acids, is responsible for the anandamide-induced

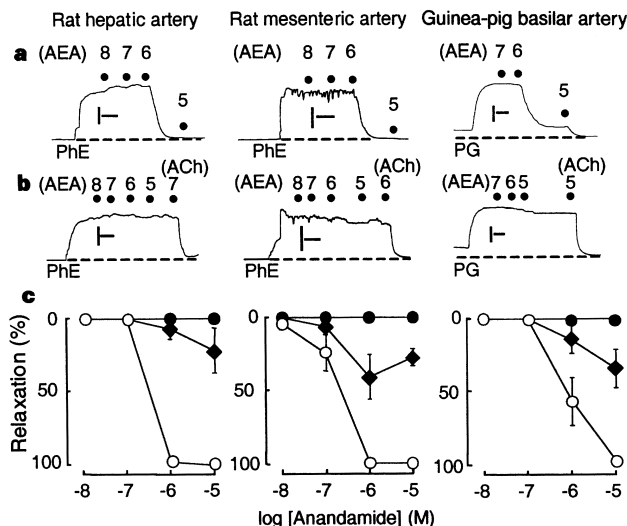


Figure 1 Effects of capsaicin and the CGRP receptor antagonist 8-37 CGRP on relaxations induced by anandamide (AEA) and acetylcholine (ACh) in rat and guinea-pig arteries. **a**, Typical relaxant responses to anandamide in these blood vessels contracted with phenylephrine (PhE; 1–3 μ M) or prostaglandin $F_{2\alpha}$ (PG; 0.3–1 μ M). Drug concentrations are given as $-\log$ molar concentrations. Horizontal and vertical scale markers represent 2 min and 2 mN, respectively. Dashed line indicates the basal tension level before addition of drugs. Each trace was obtained from a different arterial segment. **b**, Pretreatment with capsaicin (10 μ M) for 1 h (followed by washout of capsaicin for 20 min) abolished the anandamide-induced relaxation, whereas subsequent application of acetylcholine (0.1–10 μ M) always elicited complete relaxation (mediated by endothelium-derived hyperpolarising factor^{2,3,24}) in 6 out of 6 rat hepatic, 4 out of 4 rat mesenteric and 5 out of 5 guinea-pig basilar arteries. **c**, Concentration–response curves for anandamide after treatment with capsaicin (10 μ M; filled circles), 8-37 CGRP (2 μ M; diamonds) for 30 min or vehicle (circles) in the different arteries ($n = 5-8$).

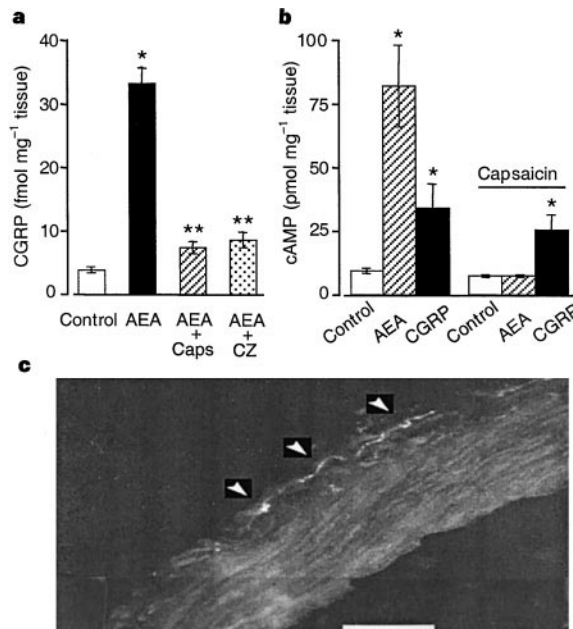


Figure 2 Release of CGRP from sensory nerves and effects of anandamide (AEA) and CGRP on cAMP levels in the rat hepatic artery. **a**, The effect of 10 min exposure to anandamide (10 μ M) on CGRP release was measured in untreated vessels ($n = 11$) and in vessels pretreated with capsaicin (Caps; 10 μ M, $n = 4$), followed by washout of capsaicin for 20 min, or incubated with capsazepine (CZ; 10 μ M, 20 min, $n = 5$). Control refers to preparations exposed to vehicle ($n = 6$). **b**, cAMP contents were measured after 5 min in the absence (control) or presence of either anandamide (AEA; 10 μ M) or CGRP (10 nM) before and after pretreatment with capsaicin as above ($n = 6$). * $P < 0.05$ compared to controls, ** $P < 0.05$ compared to anandamide alone. **c**, CGRP-like immunoreactive nerve fibres were observed in the adventitia. Bar, 40 μ m.

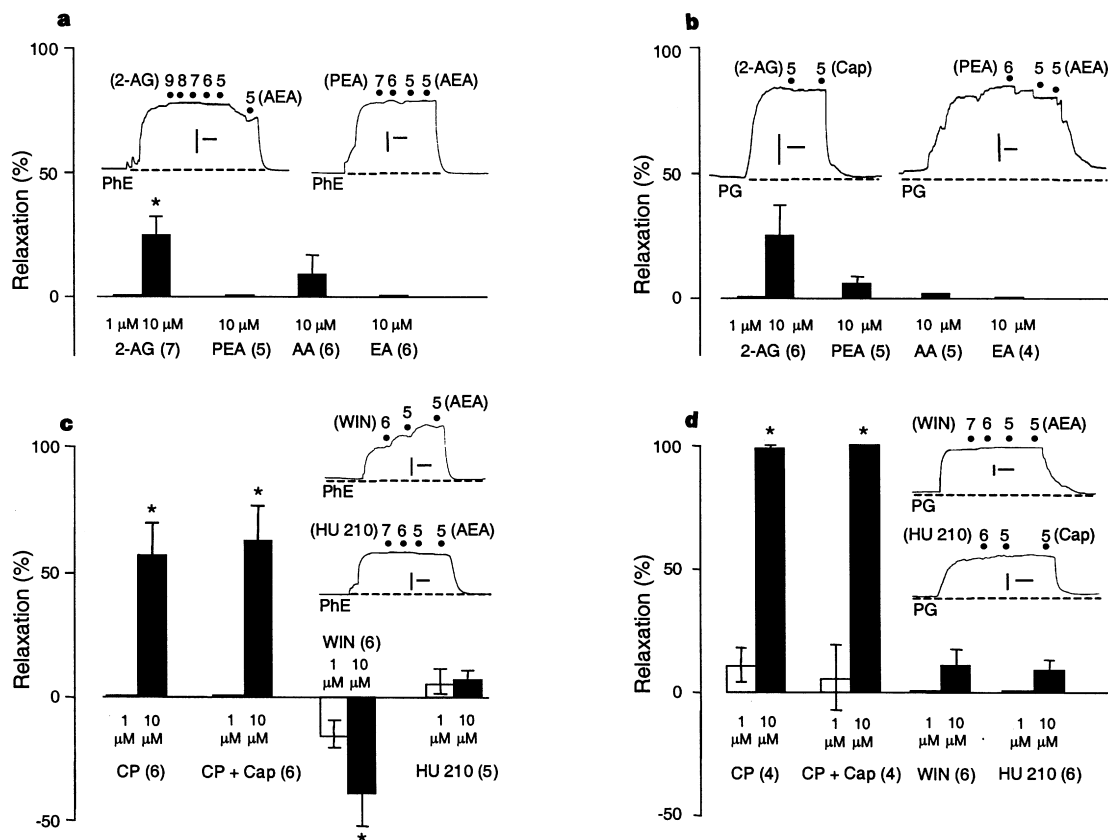


Figure 3 Relaxant effects of endogenous and synthetic cannabinoid receptor agonists in rat hepatic and guinea-pig basilar arteries. **a, b**, Effects of the endocannabinoids 2-arachidonylglycerol (2-AG) and palmitylethanolamide (PEA), and arachidonic acid (AA) and ethanolamine (EA), which are breakdown products of anandamide, in **(a)** rat hepatic and **(b)** guinea-pig basilar arteries contracted with phenylephrine (PhE; 1–3 μ M) or prostaglandin $F_{2\alpha}$ (PG; 0.1–1 μ M), respectively. Anandamide (AEA; 10 μ M) and capsaicin (Cap; 10 μ M) completely relaxed the vessels. Drug concentrations are given as –log molar concentrations. Horizontal and vertical scale markers represent 2 min and 2 mN, respectively.

relaxation in bovine coronary artery²⁰. However, this relaxation is endothelium-dependent, in contrast to the vasodilator response to anandamide in rat hepatic³ and mesenteric¹⁷ and guinea-pig basilar (unpublished observations) arteries. Cyclo-oxygenase products can be ruled out in this study, as indomethacin was present in all experiments, and eicosatrienoic acids cannot relax rat hepatic²³ and guinea-pig basilar²⁴ arteries. Furthermore, arachidonic acid (10 μ M) and ethanolamine (10 μ M) had no effect in these arteries (Fig. 3a, b), whereas methanandamide, which is resistant to enzymatic hydrolysis¹, was at least as potent as anandamide as a vasodilator in rat hepatic (methanandamide: $pEC_{50} = 6.53 \pm 0.08$, $E_{max} = 97 \pm 1\%$; $n = 8$; anandamide: $pEC_{50} = 6.49 \pm 0.01$, $E_{max} = 100 \pm 1\%$; $n = 6$) and mesenteric²¹ arteries. These findings indicate that anandamide itself is responsible for the vasodilator activity.

Capsaicin and other vanilloid compounds act by binding to specific vanilloid receptors on primary sensory neurons. Molecular studies have shown that the cloned vanilloid receptor (VR1) is a calcium-permeable, non-selective cation channel that is gated by noxious heat or extracellular protons⁸. In addition to these physiological stimuli, VR1 might also be modulated by endogenous, capsaicin-like substances, such as endocannabinoids or other lipids. Olvanil (but not capsaicin) binds to both the anandamide transporter and the CB1 receptor, indicating that capsaicin-like compounds and endocannabinoids can indeed recognize the same proteins²². VR1 is closely related to members of the transient receptor potential (TRP) channel family⁸. Membrane-derived second messengers, such as diacylglycerol or arachidonic acid,

Dashed line indicates the basal tension level before addition of drugs. Each trace was obtained from a different arterial segment. **c, d**, Among the synthetic compounds tested (HU 210, WIN 55,212-2 and CP 55,940), only CP 55,940 caused relaxation in **(c)** rat hepatic and **(d)** guinea-pig basilar arteries. In contrast to anandamide, CP 55,940 also relaxed arteries after treatment with capsaicin (10 μ M). Anandamide (10 μ M) or capsaicin (10 μ M) relaxed vessels exposed to HU 210 and WIN 55,212-2. The number of experiments is shown in brackets. * $P < 0.05$ compared to vehicle (0.2% ethanol).

can activate certain TRP-channel subtypes at micromolar concentrations^{25,26}. To determine whether anandamide interacts with native vanilloid receptors, we examined the effect of capsazepine, a competitive vanilloid receptor antagonist^{6,7}, on the vasodilator responses to anandamide, methanandamide and capsaicin. As shown in rat hepatic and mesenteric and guinea-pig basilar arteries, capsazepine significantly inhibits the vasodilator effects of anandamide and capsaicin (Fig. 4a). Capsazepine also suppresses the anandamide-induced release of CGRP from capsaicin-sensitive nerves (Fig. 2a). Further experiments on the rat hepatic artery showed that capsazepine produces a rightward parallel shift of the concentration–response curves for both capsaicin and methanandamide without affecting the maximal response, consistent with a competitive mode of inhibition (Fig. 4b, c). The Schild plots for capsazepine using capsaicin and methanandamide as agonists were not significantly different, and revealed pA_2 values (mean $\pm$ s.d.) of 6.24 ± 0.14 and 6.34 ± 0.06 , respectively (Fig. 4d), strongly favouring an identical site of action for these agonists.

To exclude the possibility that anandamide causes relaxation by activating neuronal calcium channels, we examined the effects of nimodipine and the ω -conotoxins GVIA and MVIIC, inhibitors of L-, N- and P/Q-type calcium channels²⁷, on the vasodilator response to anandamide in the rat hepatic artery. Neither ω -conotoxin GVIA (0.1 μ M) plus MVIIC (0.1 μ M) nor nimodipine (0.1 μ M) in combination with these toxins inhibited the anandamide-induced vasodilation (control: $pEC_{50} = 6.56 \pm 0.09$, $E_{max} = 98 \pm 1\%$; $n = 6$; ω -conotoxin GVIA and MVIIC: $pEC_{50} = 6.58 \pm 0.08$;

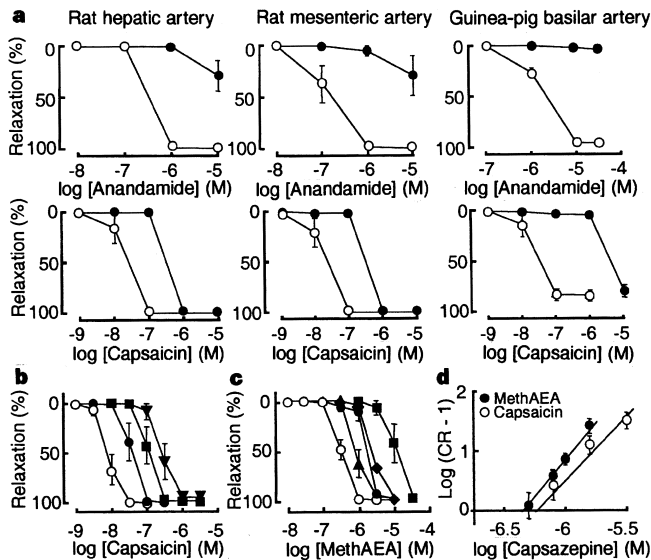


Figure 4 Effect of the vanilloid receptor antagonist capsazepine on relaxations elicited by anandamide, methanandamide and capsaicin in rat and guinea-pig arteries. **a**, Concentration-response curves for anandamide ($n = 5-6$) and capsaicin ($n = 5-6$) after incubation with capsazepine ($3 \mu\text{M}$) for 30 min (filled circles) or vehicle (open circles; 0.3% ethanol) in arteries contracted with phenylephrine (PHE; $1-3 \mu\text{M}$) or prostaglandin $\text{F}_{2\alpha}$ (PG; $0.1-1 \mu\text{M}$). **b**, **c**, Concentration-response curves for capsaicin ($n = 6$) and methanandamide (MethAEA; $n = 5-7$) in the absence (open circles) and presence of various concentrations of capsazepine ($0.5 \mu\text{M}$, upward triangles; $0.8 \mu\text{M}$, filled circles; $1.0 \mu\text{M}$, diamonds; $1.6 \mu\text{M}$, squares; $3.2 \mu\text{M}$, downward triangles) in rat hepatic arteries. **d**, Schild plots for capsazepine, using capsaicin (open circles) or methanandamide (filled circles) as agonists. For clarity, the individual data points are summarized as mean $\pm$ s.e.m. (vertical lines; $n = 5-8$). The slopes of the regression lines for capsaicin and methanandamide were (mean $\pm$ s.d.) 2.12 ± 0.36 and 2.41 ± 0.32 , respectively.

$E_{\text{max}} = 99 \pm 1\%$, $n = 6$; nimodipine and ω -conotoxin GVIA and MVIIC: $\text{pEC}_{50} = 6.31 \pm 0.02$, $E_{\text{max}} = 93 \pm 3\%$, $n = 5$). α -Latrotoxin is a neurotoxin that causes neurotransmitter release after binding to presynaptic nerve terminals²⁸. The vasodilator response to α -latrotoxin (1 nM) was unaffected by capsazepine ($3 \mu\text{M}$), but abolished by CGRP 8-37 ($2 \mu\text{M}$) or capsaicin pretreatment in the rat hepatic artery (control: $98 \pm 1\%$, $n = 12$; capsazepine: $93 \pm 2\%$, $n = 6$; CGRP 8-37: $-8 \pm 3\%$, $n = 6$; capsaicin: $-16 \pm 8\%$, $n = 7$). This indicates that capsazepine does not act through depletion of neurotransmitters from sensory nerves. Furthermore, capsazepine ($3 \mu\text{M}$) had no effect on the CGRP-induced relaxation in rat hepatic (control: $\text{pEC}_{50} = 9.46 \pm 0.01$, $E_{\text{max}} = 100 \pm 1\%$; capsazepine: $\text{pEC}_{50} = 9.39 \pm 0.04$, $E_{\text{max}} = 100 \pm 1\%$, $n = 6$), mesenteric (control: $\text{pEC}_{50} = 10.4 \pm 0.04$, $E_{\text{max}} = 99 \pm 1\%$; capsazepine: $\text{pEC}_{50} = 10.3 \pm 0.12$, $E_{\text{max}} = 99 \pm 1\%$, $n = 4$) and guinea-pig basilar (control: $\text{pEC}_{50} = 8.72 \pm 0.16$, $E_{\text{max}} = 98 \pm 1\%$; capsazepine: $\text{pEC}_{50} = 8.70 \pm 0.15$, $E_{\text{max}} = 99 \pm 1\%$, $n = 4-5$) arteries, confirming the specificity of capsazepine.

To obtain direct evidence that anandamide interacts with vanilloid receptors, we examined the effects of anandamide and other lipid messengers on the cloned VR1 channel. Whole-cell patch-clamp recordings from VR1-transfected HEK293 cells showed robust membrane currents when anandamide was applied by bath perfusion (Fig. 5a, top left); such currents were not observed in vector-transfected HEK293 cells ($n = 5$; data not shown). Similar responses were seen with inside-out membrane patches excised from VR1-expressing *Xenopus* oocytes (Fig. 5a, top, middle), but not uninjected oocytes ($n = 4$; not shown). In each case, anandamide elicited outwardly rectifying cationic currents characteristic of VR1 (Fig. 5a, bottom). In VR1-transfected HEK293 cells, pEC_{50} for anandamide was estimated as 5.31 ± 0.06 ($n = 13$). Thus, anandamide

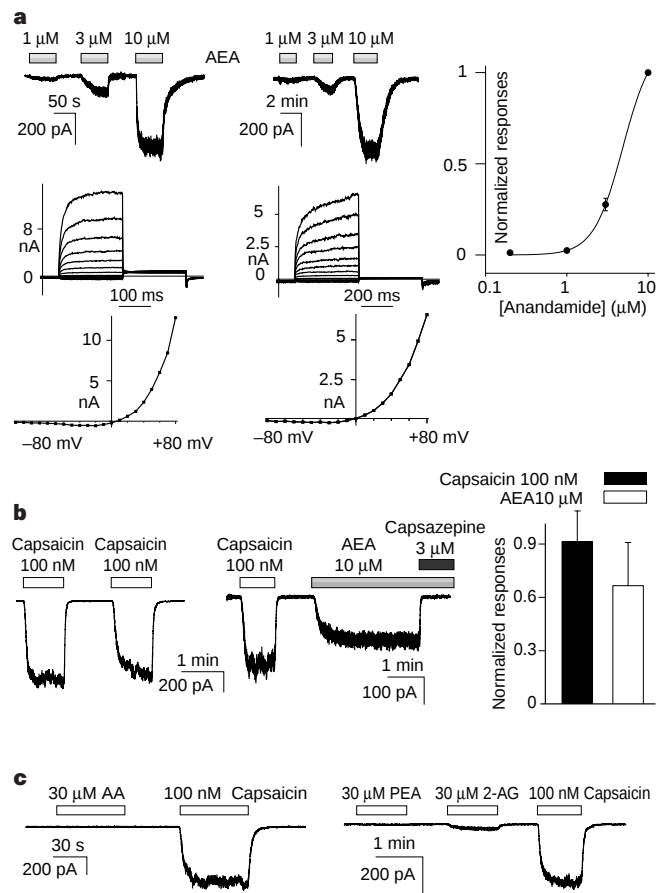


Figure 5 Anandamide activates the cloned vanilloid receptor (VR1). **a**, Patch-clamp recordings from VR1-expressing HEK293 cells (top left, whole-cell configuration) or *Xenopus* oocytes (top middle, excised inside-out membrane patch) show concentration-dependent responses to bath-applied anandamide (plotted for VR1-transfected HEK293 cells, top right; $n = 6-24$). The curve fit revealed a Hill coefficient for anandamide of 2.41 ± 0.46 , which is close to that obtained for capsaicin in VR1-transfected *Xenopus* oocytes (2.08) (ref. 8). Current-voltage relationships for responses in HEK293 cells or oocytes (bottom left and right, respectively) showed pronounced outward rectification characteristic of VR1 in the presence of anandamide ($10 \mu\text{M}$). Currents were generated by stepping from a holding potential of -80 mV to the test potentials as indicated in the figure. **b**, Relative responses of VR1 to capsaicin ($n = 7$) and anandamide (AEA; $n = 11$) were determined by whole-cell patch-clamp analysis of VR1-expressing HEK293 cells. Currents were normalized to an initial response to 100 nM capsaicin. Anandamide-evoked currents were inhibited by co-application of $3 \mu\text{M}$ capsazepine ($n = 8$). **c**, VR1-expressing HEK293 cells did not respond to arachidonic acid (AA; $n = 8$) or palmitylethanolamide (PEA; $n = 9$), whereas small membrane currents were elicited by 2-arachidonylglycerol (2-AG; $n = 9$).

is at least as potent on VR1 as other lipid messengers on *Drosophila* or mammalian TRP channels^{25,26}. At a near maximal concentration ($10 \mu\text{M}$), anandamide evoked a response similar to that obtained with 100 nM capsaicin (Fig. 5b). Relative to capsaicin, the efficacy of anandamide as a VR1 agonist was estimated as 30–50%. Anandamide ($10 \mu\text{M}$) was unable to induce membrane currents in HEK293 cells expressing cloned serotonin-gated ($5\text{-HT}_3\text{R}$ -A, $n = 6$; data not shown) or ATP-gated (P_2X_2 , $n = 7$; data not shown) ion channels, whereas these transfected cells consistently responded to $10 \mu\text{M}$ serotonin (current range, $0.5-2 \text{ nA}$) or $10 \mu\text{M}$ ATP (current range, $4-7 \text{ nA}$), providing additional proof that anandamide interacts specifically with VR1. The observation that capsazepine completely and rapidly antagonized anandamide-induced currents further supports this conclusion (Fig. 5b). In VR1-expressing HEK293 cells, neither arachidonic acid nor

palmitylethanolamide evoked significant inward current responses, and 2-arachidonylethanolamide elicited only a small current when applied at a concentration of 30 μ M (Fig. 5c). Diacylglycerol activates human TRPC3 and TRPC6, but not rat VR1 channels²⁶. Together, these observations show that anandamide, but not all lipids, can activate VR1.

Although capsaicin was a more potent agonist than anandamide or methanandamide in all our test systems, the absolute potencies of these compounds were always higher (3–30-fold) in the vascular preparations than in VR1-expressing HEK293 cells and *Xenopus* oocytes^{8,29}. The reason for the difference in sensitivity to vanilloid receptor agonists between these test systems is unclear, but obvious possibilities are the presence in intact tissues of large receptor reserves, effective cellular amplification systems and/or potential VR1 channel regulators, which are absent in heterologous expression systems. Such differences may also account for the observation that, although methanandamide and anandamide are equally effective as vasodilators, methanandamide is only about 10% as effective as anandamide in producing membrane currents in HEK293 cells expressing VR1 (data not shown).

Our findings indicate that anandamide activates vanilloid receptors on primary sensory neurons, and provide a likely molecular mechanism for the non-CB1 receptor-mediated vasodilator action of anandamide. This raises the possibility that anandamide and/or structurally-related lipids act as endogenous vanilloid receptor modulators and thereby participate in the regulation of the various afferent (nociception, visceral reflexes) and efferent (for example local vasodilation and neurogenic inflammation) functions subserved by capsaicin-sensitive sensory neurons. □

Methods

Recording of tension. Experiments were performed on hepatic (400 μ m outer diameter) and mesenteric (200 μ m outer diameter) arteries from female Wistar-Hannover rats (200–250 g) and on basilar arteries (300 μ m outer diameter) from male guinea-pigs (300 g). The arteries were cut into ring segments and mounted in tissue baths containing aerated physiological salt solution (5% CO₂ and 95% O₂; 37 °C; pH 7.4) as described^{3,24}. We carried out all experiments in the presence of N^G-nitro-L-arginine (0.3 mM) and indomethacin (10 μ M) to eliminate any contribution of nitric oxide and cyclooxygenase products, respectively. We studied relaxant responses in preparations contracted with phenylephrine (rat hepatic and mesenteric arteries) or prostaglandin F_{2 α} (guinea-pig basilar artery)^{3,24}. When stable contractions were obtained, agonists were added cumulatively to determine concentration–response relationships.

Measurements of CGRP and cyclic AMP. Segments of rat hepatic arteries were equilibrated for 1 h in aerated physiological salt solution (5% CO₂ and 95% O₂; 37 °C; pH 7.4), containing N^G-nitro-L-arginine (0.3 mM) and indomethacin (10 μ M). CGRP: preparations were transferred to eppendorf tubes, containing physiological salt solution with the addition of 0.05% BSA and the test drugs. The segments were removed after 5 min, and the solution in the test tubes was evaporated. We determined the amount of CGRP in the pellet using rat ¹²⁵I-CGRP radioimmunoassay RIA kit (Peninsula labs). The amount of cyclic AMP after 5 min exposure to the test drugs was determined as described³⁰.

Immunohistochemistry. Rat hepatic arteries were fixed in 4% formaldehyde for 4 h, rinsed in a phosphate buffer (pH 7.4) containing 15% sucrose for three days and then frozen. Cryostat sections were cut at a thickness of 10 μ m and prepared for reaction with CGRP antiserum (EURO-DIAGNOSTICA, Malmö, Sweden) at 1:320 dilution for 24 h (room temperature). CGRP-like immunoreactivity was visualized after exposure to FITC-conjugated goat anti-guinea-pig IgG (Sigma, USA) at 1:80 dilution for 90 min.

Patch-clamp. HEK293 cells were transfected with 0.5 μ g VR1 cDNA, using lipofectamine (Gibco), and plated on glass coverslips 24–72 h before analysis. *Xenopus* oocytes were prepared as described⁸ and injected with 2.5 ng of VR1 cRNA. Standard bath solution for all recordings contained (in mM): 10 Tris, 150 CsCl, 1 MgCl₂, 1 EGTA, pH 7.4. For oocyte patch recording, this solution was also included in the pipette, whereas for whole-cell recording from HEK293 cells, the pipette contained a 0.9 times dilution of this solution. For

all continuous recordings, the membrane was held at –40 mV. All experiments were done at room temperature. Drugs were prepared as ethanol stocks and diluted with bath solution, containing 0.2 mM beta-cyclodextrin, and this solution was bath sonicated for 5 min in ice water before use. Agonist washout was achieved with bath solution, containing 0.2 mM beta-cyclodextrin.

Calculations and statistics. Relaxations are expressed as percentage reversal of the agonist-induced contraction (100% indicates a complete relaxation). We calculated drug concentrations eliciting 50% (EC₅₀) and maximal (E_{max}) relaxation as described²⁴. Data are presented as means $\pm$ s.e.m. (vertical lines in Figures), and *n* indicates the number of vascular segments (animals) or cells examined. pA₂ and slope values of the Schild plots were estimated according to the Jackknife method. Statistical analysis was performed by using Student's *t*-test, Mann–Whitney *U*-test or analysis of variance (ANOVA), followed by Bonferroni Dunn's post hoc test (Statview 4.12). Cyclic AMP values were log transformed. Statistical significance was accepted when *P* < 0.05.

Drugs. Anandamide, R(+)-methanandamide, 2-arachidonylethanolamide, WIN 55,212-2 (RBI); HU 210, CP 55,940 (Tocris); palmitylethanolamide (Biomol); capsaicin, capsaizipine (Fluka); SR141716A (Sanofi Winthrop); arachidonic acid (Sigma) were all dissolved in ethanol. Distilled water was used as solvent for acetylcholine chloride, L-phenylephrine hydrochloride, N^G-nitro-L-arginine, human 8-37 CGRP, ethanolamine hydrochloride (Sigma); human α -CGRP (Peninsula); α -latrotoxin, ω -conotoxins GVIA and MVIIC (Alomone); prostaglandin F_{2 α} (Upjohn); indomethacin (Confortid, Dumex). Nimodipine (Nimotop, Bayer) was diluted with saline.

Received 18 March; accepted 21 June 1999.

- Pertwee, R. G. Pharmacology of cannabinoid CB₁ and CB₂ receptors. *Pharmacol. Ther.* **74**, 129–180 (1997).
- Plane, F., Holland, M., Waldron, G. J., Garland, C. J. & Boyle, J. P. Evidence that anandamide and EDHF act via different mechanisms in rat isolated mesenteric arteries. *Br. J. Pharmacol.* **121**, 1509–1511 (1997).
- Zygmunt, P. M. *et al.* Studies on the effects of anandamide in rat hepatic artery. *Br. J. Pharmacol.* **122**, 1679–1686 (1997).
- Wagner, J. A., Varga, K., Jarai, Z. & Kunos, G. Mesenteric vasodilation mediated by endothelial anandamide receptors. *Hypertension* **33**, 429–434 (1999).
- Han, S.-P., Naes, L. & Westfall, T. C. Inhibition of periaortic nerve stimulation-induced vasodilation of the mesenteric arterial bed by CGRP (8-37) and CGRP receptor desensitization. *Biochem. Biophys. Res. Commun.* **168**, 786–791 (1990).
- Bevan, S. *et al.* Capsazepine: a competitive antagonist of the sensory neurone excitant capsaicin. *Br. J. Pharmacol.* **107**, 544–552 (1992).
- Szallasi, A. & Blumberg, P. M. Vanilloid (capsaicin) receptors and mechanisms. *Pharmacol. Rev.* **51**, 159–212 (1999).
- Caterina, M. J. *et al.* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**, 816–824 (1997).
- Devane, W. A. *et al.* Isolation and structure of a brain constituent that binds to the cannabinoid receptor. *Science* **258**, 1946–1949 (1992).
- Deutsch, D. G. *et al.* Production and physiological actions of anandamide in the vasculature of the rat kidney. *J. Clin. Invest.* **100**, 1538–1546 (1997).
- Di Marzo, V., De Petrocellis, L., Sepe, N. & Buono, A. Biosynthesis of anandamide and related acylethanolamides in mouse J774 macrophages and N₁₈ neuroblastoma cells. *Biochem. J.* **316**, 977–984 (1996).
- Varga, K., Lake, K. D., Huangfu, D., Guyenet, P. G. & Kunos, G. Mechanism of the hypotensive action of anandamide in anesthetized rats. *Hypertension* **28**, 682–686 (1996).
- Wagner, J. A. *et al.* Activation of peripheral CB₁ cannabinoid receptors in haemorrhagic shock. *Nature* **390**, 518–521 (1997).
- Varga, K., Wagner, J. A., Bridgen, D. T. & Kunos, G. Platelet- and macrophage-derived endogenous cannabinoids are involved in endotoxin-induced hypotension. *FASEB J.* **12**, 1035–1044 (1998).
- Ishac, E. J. N. *et al.* Inhibition of exocytotic noradrenaline release by presynaptic cannabinoid CB₁ receptors on peripheral sympathetic nerves. *Br. J. Pharmacol.* **118**, 2023–2028 (1996).
- Sugiura, T. *et al.* Detection of an endogenous cannabinimimetic molecule 2-arachidonylethanolamide, and cannabinoid CB₁ receptor mRNA in human vascular cells: is 2-arachidonylethanolamide a possible vasomodulator? *Biochem. Biophys. Res. Commun.* **243**, 838–843 (1998).
- White, R. & Hiley, C. R. The actions of some cannabinoid receptor ligands in the rat isolated mesenteric artery. *Br. J. Pharmacol.* **125**, 533–541 (1998).
- Kawasaki, H., Takasaki, K., Saito, A. & Goto, K. Calcitonin gene-related peptide acts as a novel vasodilator neurotransmitter in mesenteric resistance vessels of the rat. *Nature* **335**, 164–167 (1988).
- Edvinsson, L., Ekman, R., Jansen, I., McCulloch, J. & Uddman, R. Calcitonin gene-related peptide and cerebral blood vessels: distribution and vasomotor effects. *J. Cereb. Blood Flow Metab.* **7**, 720–728 (1987).
- Pratt, P. F., Hillard, C. J., Edgemond, W. S. & Campbell, W. B. N-arachidonyl ethanolamide relaxation of bovine coronary artery is not mediated by CB₁ cannabinoid receptor. *Am. J. Physiol.* **274**, H375–H381 (1998).
- White, R. & Hiley, C. R. The actions of the cannabinoid receptor antagonist, SR 141716A, in the rat isolated mesenteric artery. *Br. J. Pharmacol.* **125**, 689–696 (1998).
- Di Marzo, V. *et al.* Interactions between synthetic vanilloids and the endogenous cannabinoid system. *FEBS Lett.* **436**, 449–454 (1998).
- Zygmunt, P. M., Edwards, G., Weston, A. H., Davis, S. C. & Högestätt, E. D. Effects of cytochrome P450 inhibitors on EDHF-mediated relaxation in the rat hepatic artery. *Br. J. Pharmacol.* **118**, 1147–1152 (1996).
- Petersson, J., Zygmunt, P. M., Jönsson, P. & Högestätt, E. D. Characterization of endothelium-dependent relaxation in guinea pig basilar artery—effects of hypoxia and role of cytochrome P₄₅₀ mono-oxygenase. *J. Vasc. Res.* **35**, 285–294 (1998).
- Chyb, S., Raghu, P. & Hardie, R. Polyunsaturated fatty acids activate the drosophila light-sensitive channels TRP and TRPL. *Nature* **397**, 255–259 (1999).

26. Hofmann, T. *et al.* Direct activation of human TRPC6 and TRPC3 channels by diacylglycerol. *Nature* **397**, 259–263 (1999).
27. Olivera, B. M., Miljanich, G. P., Ramachandran, J. & Adams, M. E. Calcium channel diversity and neurotransmitter release: the omega-conotoxins and omega-agatoxins. *Annu. Rev. Biochem.* **63**, 823–867 (1994).
28. Geppert, M. *et al.* Neurexin Ia is a major α -latrotoxin receptor that cooperates in α -latrotoxin action. *J. Biol. Chem.* **273**, 1705–1710 (1998).
29. Tominaga, M. *et al.* The cloned capsaicin receptor integrates multiple pain-producing stimuli. *Neuron* **21**, 531–543 (1998).
30. Zygmunt, P. M., Grundemar, L. & Högestätt, E. D. Endothelium-dependent relaxation resistant to N ω -nitro-L-arginine in the rat hepatic artery and aorta. *Acta Physiol. Scand.* **152**, 107–114 (1994).

Acknowledgements. This work was supported by the Swedish Medical Research Council, the Medical Faculty of Lund (ALF) and the Crafoords Foundation (E.D.H.), the National Institutes of Health and the McKnight Foundation for Neuroscience (H.C. and D.J.) and the Ministero dell'Università e della Ricerca Scientifica (V.D.M.). P.M.Z. was supported by the Swedish Society for Medical Research and the Swedish Medical Research Council. J.P. was supported by the Medical Faculty of Lund (ALF).

Correspondence and requests for materials should be addressed to E.D.H. (e-mail: Edward.Hogestatt@klinfarm.lu.se).

Munc13-1 is essential for fusion competence of glutamatergic synaptic vesicles

Iris Augustin^{*}, Christian Rosenmund[†], Thomas C. Südhof[‡] & Nils Brose^{*}

^{*}Max-Planck-Institut für Experimentelle Medizin, AG Molekulare Neurobiologie, Hermann-Rein-Str. 3, D-37075 Göttingen, Bundesrepublik Deutschland

[†]Max-Planck-Institut für Biophysikalische Chemie, Abteilung Membranbiophysik, Am Fassberg 11, D-37077 Göttingen, Bundesrepublik Deutschland

[‡]Center for Basic Neuroscience, Department of Molecular Genetics, Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, Texas 75235, USA

Neurotransmitter release at synapses between nerve cells is mediated by calcium-triggered exocytotic fusion of synaptic vesicles¹. Before fusion, vesicles dock at the presynaptic release site where they mature to a fusion-competent state^{1,2}. Here we identify Munc13-1, a brain-specific presynaptic phorbol ester receptor^{3,4}, as an essential protein for synaptic vesicle maturation. We show that glutamatergic hippocampal neurons from mice lacking Munc13-1 form ultrastructurally normal synapses whose synaptic-vesicle cycle is arrested at the maturation step. Transmitter release from mutant synapses cannot be triggered by action potentials, calcium-ionophores or hypertonic sucrose solution. In contrast, release evoked by α -latrotoxin is indistinguishable from wild-type controls, indicating that the toxin can bypass Munc13-1-mediated vesicle maturation. A small subpopulation of synapses of any given glutamatergic neuron as well as all synapses of GABA (γ -aminobutyric acid)-containing neurons are unaffected by Munc13-1 loss, demonstrating the existence of multiple and transmitter-specific synaptic vesicle maturation processes in synapses.

Deletion mutations in the mouse gene for Munc13-1 were generated by homologous recombination in embryonic stem cells. Munc13-1-deficient mice do not feed and are distinguishable from wild-type animals shortly after birth by a lack of milk in their stomachs, weakness and reduced breathing rate. Mutant mice die within a few hours of birth. Anatomical and morphological analyses of newborn offspring from heterozygous interbreedings showed that the overall structure and cytoarchitecture of mutant brains and the distribution of synaptic markers (synaptotagmin, synapsin) were indistinguishable from those of wild-type controls (data not shown). Immunoblotting of brains from newborn littermates showed that, in addition to Munc13-1, which was completely absent in mutant mice, 4 out of 35 pre- and postsynaptic proteins

tested showed small changes (complexin I, $82 \pm 8\%$; Rab3A, $76 \pm 9\%$; Munc13-1-interactor syntaxin IA/B, $85 \pm 5\%$; synaptobrevin/VAMP II, $128 \pm 6\%$ of wild-type control). In contrast, the levels of other Munc13 isoforms (Munc13-2 and Munc13-3) were not changed in mutant brains (see Supplementary Information).

To determine the role of Munc13-1 in synaptic transmission, we performed electrophysiological analyses of cultured hippocampal neurons from newborn pups. Routinely, recordings were done on single hippocampal neurons that were cultured on microdot islands where they form recurrent (autaptic) synapses onto themselves. Cultures obtained from wild-type and mutant hippocampi did not differ in overall morphology and cell density. However, whole-cell patch-clamp analysis revealed a marked phenotypical change in Munc13-1-deficient cultures. Although wild-type and heterozygous hippocampal neurons showed similar robust excitatory postsynaptic responses upon evoked release (1.43 ± 0.43 nA for +/+, $n = 50$; 1.58 ± 0.25 nA for +/-, $n = 48$), synaptic responses in mutant cells were reduced by more than 90% (0.11 ± 0.01 nA, $n = 134$; Fig. 1a, c). This was not due to a change in the density of voltage-gated conductances, as mutant cells generated normal action potentials in response to current injections, and their sodium, potassium and calcium current densities were not significantly different from normal (not shown). Inhibitory GABA-mediated transmission measured in cells from the same cultures appeared to be unaffected by the lack of Munc13-1 (1.93 ± 0.56 nA in +/+, $n = 11$; 1.91 ± 0.19 nA in +/-, $n = 59$; versus 1.85 ± 0.14 nA in -/-, $n = 74$; Fig. 1b, c). To test whether the abnormalities in glutamatergic transmission in Munc13-1-deficient cells are caused by a selective deficiency in, or lack of, voltage-gated calcium channels in mutant synapses, we used the calcium ionophore calcimycin to bypass presynaptic voltage-gated calcium channels. In wild-type neurons, addition of calcimycin triggered robust miniature excitatory postsynaptic current (mEPSC) bursts (postsynaptic peak amplitude 0.374 ± 0.083 nA, $n = 7$) that were abolished by the glutamate receptor blocker NBQX (2,3-dihydroxy-6-nitro-7-sulphamoylbenzene quinoxaline). In contrast, Munc13-1-deficient cells responded weakly (0.017 ± 0.007 nA, $n = 9$) (Fig. 1d, e), showing that changes in calcium influx cannot explain the mutant phenotype. The observed specific reduction in glutamatergic transmission was not due to loss of synapses, as the density of total synaptic contacts (not shown) and the density of glutamatergic synapses were not altered by the mutation (Fig. 1f).

Given that synapses form at normal densities in Munc13-1-deficient neuronal cultures (Fig. 1f) and that mutant neurons generate normal action potentials, the marked and specific reduction in evoked excitatory synaptic responses resulting from loss of Munc13-1 could be caused by either a defect in the presynaptic release machinery, or an alteration in postsynaptic sensitivity. From the following results, we conclude that the postsynaptic glutamate-receptor sensitivity of hippocampal neurons was not affected by the Munc13-1 mutation: mEPSC amplitudes and their probability density were not changed in mutant cells as compared to wild-type controls (14.6 ± 1.2 pA, $n = 465$, five -/- cells versus 15.1 ± 1.5 pA, $n = 523$, five +/- cells), and responses to exogenous application of glutamatergic agonists were identical in mutant and wild-type neurons. Responses evoked by 30 μ M kainate resulted in inward currents of 1.47 ± 0.4 nA ($n = 6$) and 1.87 ± 0.69 nA ($n = 6$) for wild-type and -/- neurons, respectively. NMDA (N-methyl-D-aspartate) responses (10 μ M) were 2.02 ± 0.31 nA (wild type, $n = 5$) versus 1.75 ± 0.23 (-/-, $n = 12$). Furthermore, expression of the postsynaptic NMDA receptor isoform NMDA R1 was not altered in Munc13-1 mutant brains (see Supplementary Information).

Because mutant neurons showed normal action potentials and postsynaptic sensitivity, and because Munc13-1 is specifically localized to presynaptic active zones in rat hippocampus⁴, presynaptic mechanisms are the most likely cause for the aberrant phenotype of Munc13-1-deficient neurons. In principle, these mechanisms

include the membrane-trafficking processes of the synaptic vesicle cycle *per se*, the transmitter loading of vesicles and the triggering of the exocytotic fusion reaction by calcium. Our experiments with the calcium ionophore calcimycin showed that deficiencies in synaptic voltage-gated calcium channels do not cause the Munc13-1-mutant phenotype (Fig. 1d, e). Likewise, loading of synaptic vesicles with glutamate appeared to be unaffected by the mutation, as mEPSC amplitudes were similar in wild-type and knockout neurons, indicating that similar amounts of glutamate are released upon fusion of a single vesicle (see above). In view of these findings, we examined the remaining possibility, that loss of Munc13-1 might interfere directly with membrane-trafficking in the synaptic-vesicle cycle. We started by quantifying the pool of readily releasable synaptic vesicles (mature and fusion-competent vesicles) in wild-type and mutant neurons. The readily releasable vesicle pool is defined as those quanta that are released during the transient burst of exocytotic activity following application of hypertonic sucrose solution⁵. It is the same as the pool drawn upon when release is activated by action potentials⁶. We found that excitatory post-synaptic responses induced by application of hypertonic sucrose solution were markedly decreased in Munc13-1-mutant neurons as compared to control cells (Fig. 2a, b). The degree of this reduction was similar to that observed for synaptically evoked responses (Fig. 1c), and was paralleled by a concomitant reduction in mEPSC rate (Fig. 2c). GABA-mediated transmission evoked by hypertonic sucrose solution was unaltered in mutant neurons (data not shown). Thus, loss of Munc13-1 leads to a marked and selective reduction in the readily releasable vesicle pool in excitatory synapses which, in turn, causes a parallel reduction in synaptically evoked transmission.

The probability that a quantum is released at an active zone is thought to be proportional to the size of the readily releasable vesicle pool⁶⁻⁸. Therefore, it is likely that the observed reduction in the readily releasable vesicle pool in Munc13-1-deficient neurons (Fig. 2a, b) is paralleled by a reduction in the probability of synaptic release. To estimate the release probability in Munc13-1-deficient cells, we progressively blocked NMDA-receptor-mediated synaptic currents using the irreversible open-channel blocker MK-801 (refs

9, 10). The rate of progressive block by MK-801 reflects the release probability across all synapses of a given neuron, which is non-uniform in hippocampal neurons (range 0.09–0.54)^{9,10}. We found that the time constants of progressive MK-801 block in Munc13-1-deficient hippocampal neurons were indistinguishable from those observed in wild-type cells (Fig. 3a, b), indicating identical release probabilities.

Clearly, Munc13-1-mutant cells release transmitter with a normal probability despite a marked reduction in the pool of readily releasable vesicles. The only explanation for this apparent contradiction is that, in Munc13-1-deficient neurons, most synapses are silent owing to loss of Munc13-1, but a small number of synapses is unaffected by the mutation and exhibits normal transmitter-release

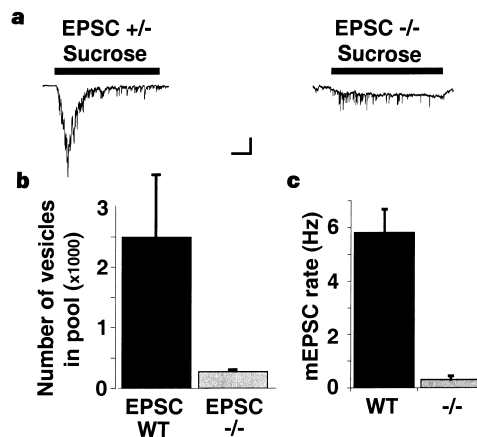


Figure 2 Impaired glutamatergic neurotransmission results from a decreased readily releasable vesicle pool. **a**, Vesicle pool measurements by application of hypertonic solution (500 mM sucrose added to the external medium, black bar) to the entire dendritic tree of a wild-type neuron and a Munc13-1-deficient neuron. Scale bar, 100 pA, 1 s. **b**, Histogram summary of the mean sizes of the readily releasable vesicle pool (expressed as number of vesicles) for glutamatergic cells (EPSC; WT, $n = 15$; $-/-$, $n = 19$; $P < 0.01$). **c**, mEPSC rate in Munc13-1-deficient neurons ($-/-$; $n = 5$) and wild-type neurons (WT; $n = 4$; $P < 0.001$).

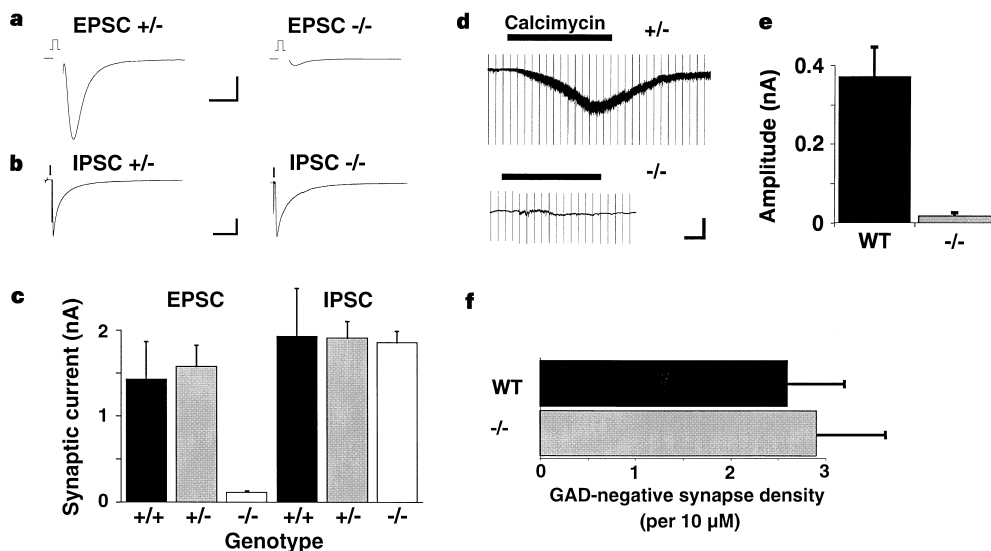


Figure 1 Glutamatergic but not GABA-mediated neurons show strongly impaired synaptic transmission in Munc13-1-deficient mice. **a**, AMPA-receptor-mediated EPSCs from a heterozygous (+/-) and a Munc13-1-deficient (-/-) glutamatergic neuron. Scale bar, 0.5 nA, 10 ms. **b**, GABA_A-mediated IPSCs from a heterozygous (+/-) and a Munc13-1-deficient (-/-) GABA-containing neuron. Scale bar, 1 nA, 50 ms. **c**, Summary of average synaptic peak current amplitudes. Data from control (+/+) and heterozygous (+/-) animals were pooled in all following experiments and classified as wild-type (WT) unless indicated otherwise.

d, Whole-cell recordings from glutamatergic mouse hippocampal neurons after treatment with the Ca²⁺-ionophore calcimycin (10 μM for 1–2 min). Vertical lines at 8-s intervals represent stimulation artefacts resulting from intermittent monitoring of evoked synaptic responses. Scale bar, 0.5 nA, 20 s. **e**, Summary histograms of peak amplitudes generated by calcimycin and blockable by NBQX (WT, $n = 7$; $-/-$, $n = 9$; $P < 0.001$). **f**, Synapsin-positive/GAD-65-negative axon varicosities per μm dendrite in wild-type (WT, +/+) and Munc13-1-deficient cultures (-/-).

characteristics. To test this directly, we again used the irreversible block of NMDA receptors by MK-801. In hippocampal neurons only very few NMDA receptors are expressed extrasynaptically: 80% of receptors are on postsynaptic sites¹¹. Therefore, complete block of NMDA receptor-mediated EPSCs by MK-801 in wild-type cells by repetitive synaptic activation should lead to an irreversible block of most cellular NMDA receptors. In contrast to this, stimulation under the same conditions of a mutant cell with mostly silent synapses should result in a much larger number of NMDA receptors remaining unaffected by the MK-801 block. Given similar synapse densities (Fig. 1f) and NMDA-receptor distributions (protein levels of NMDA R1, sensitivity to exogenous NMDA and ratio of AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) mediated/NMDA-mediated EPSCs are normal in mutant neurons; data not shown), the relative number of silent synapses can be

assayed by measuring the cellular response to exogenously applied NMDA before and after complete block of synaptically activated NMDA receptors by repetitive stimulation in the presence of MK-801. Using this paradigm, we found that, in Munc13-1-deficient neurons, only 20% of cellular NMDA receptors are activated by synaptic transmitter release and blockable by MK-801 (Fig. 3c (right), d). In contrast, 75% of cellular NMDA receptors are activated in wild-type neurons (Fig. 3c (left), d). This shows that, in Munc13-1-mutant cells, at least 70% of all synapses do not release neurotransmitter under our stimulation conditions. Clearly, a given excitatory neuron forms a heterogeneous population of synapses. Some of these are independent of Munc13-1, whereas, for most, Munc13-1 is essential.

Our electrophysiological analyses showed that the overall reduction of the readily releasable vesicle pool in Munc13-1-deficient hippocampal neurons is due to the fact that loss of Munc13-1 leads to a functional shutdown of most synapses in a mutant neuron. In principle, this mutant presynaptic phenotype could be caused by perturbations in almost any step of the synaptic-vesicle cycle. Only the last, calcium-triggered step of exocytosis represents an exception because interference of the Munc13-1 mutation with the calcium sensor or with the coupling between calcium channels and the release machinery would not lead to alterations of the readily releasable vesicle pool¹². The facts that calcium ionophores cannot bypass the Munc13-1-mutant phenotype (Fig. 1d, e) and that Munc13-1 does not bind calcium and can therefore not serve as a calcium sensor itself (A. Betz and N.B., unpublished observations) provide further evidence indicating that the Munc13-1 mutation does not affect the calcium-triggering step of exocytosis.

Changes in presynaptic membrane trafficking can result in changes in presynaptic morphology. Therefore, we performed a detailed ultrastructural analysis of wild-type and Munc13-1-deficient synapses. We studied cultured hippocampal neurons after they had been characterized electrophysiologically, and concentrated our analysis on asymmetric/type I synapses, which are glutamatergic. We obtained comparative morphometric data on the size of active zones, the number of docked vesicles per micrometre of active zone and the size of the synaptic-vesicle pool close to the active zone. We found that the ultrastructure of Munc13-1-deficient synapses was very similar to that of wild-type controls (Fig. 4). This lack of morphological changes in Munc13-1-deficient glutamatergic synapses despite a markedly decreased readily releasable vesicle

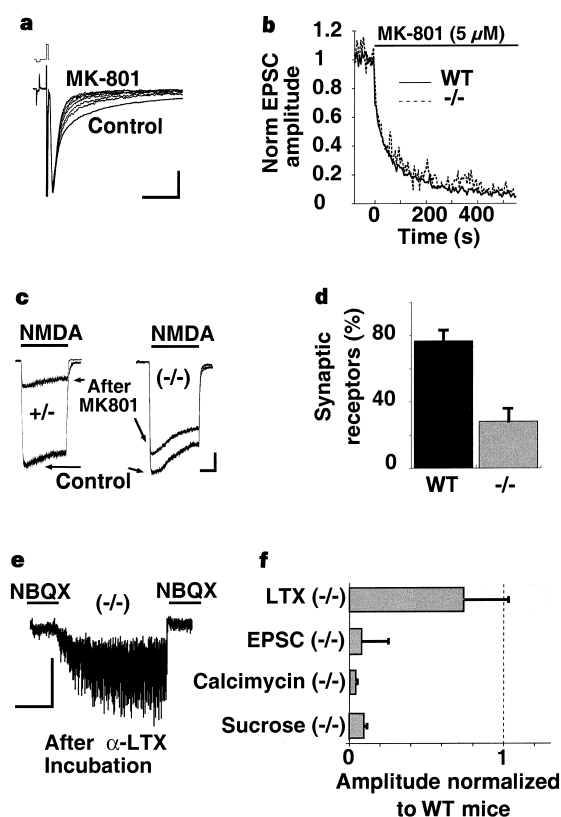


Figure 3 Selective silencing of most glutamatergic synapses in Munc13-1-deficient neurons and effect of α -latrotoxin. **a**, Dual-component EPSCs recorded from a wild-type neuron in the absence of MK801 (Control) and at episode 1, 3, 5, 10, 15, 20 and 30 following switch to MK-801-containing ($5 \mu\text{M}$) external solution. Scale bar, 1 nA, 50 ms. **b**, Time course of NMDA-mediated EPSC amplitude (mean amplitude normalized to the control period) from wild-type neurons (WT; solid line; $n = 13$) and Munc13-1-deficient neurons ($-/-$; dotted line; $n = 23$) before and following switch to $5 \mu\text{M}$ MK-801-containing external solution (black bar). Release was evoked at 0.2 Hz. **c**, Whole-cell NMDA-receptor activity after exogenous application of $10 \mu\text{M}$ NMDA in wild-type (WT, left traces) and in Munc13-1-deficient neurons ($-/-$, right traces) before (Control) and after MK-801-block of synaptically activated NMDA receptors. Scale bar, 0.5 nA, 1 s. **d**, Fraction of NMDA receptors that can be activated synaptically (WT; black bar; $n = 7$; $-/-$; grey bar; $n = 10$). **e**, Response of a glutamatergic neuron (blockable by NBQX) in a Munc13-1-deficient culture to treatment with 0.3 nM α -latrotoxin for 30 s (4 min after toxin application). Scale bar, 100 pA, 5 s. **f**, Comparison of release induced by α -latrotoxin, synaptic release (data from Fig. 1c), and release induced by hyper-tonic sucrose solution (data from Fig. 2b) or calcimycin (data from Fig. 1e). Histograms represent the response of Munc13-1-deficient glutamatergic neurons normalized to the wild-type response (grey boxes indicate standard error in wild type).

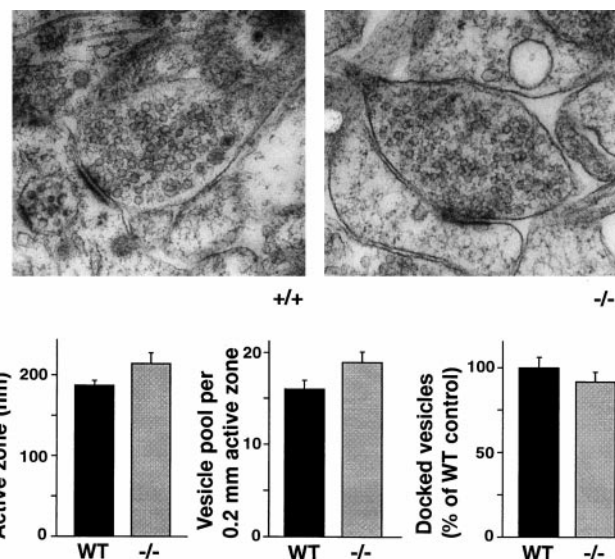


Figure 4 Munc13-1-deficient neurons form ultrastructurally normal synapses. Hippocampal neurons were analysed after 11 days in culture (122 synapses from four wild-type animals, and 54 synapses from two knockout animals).

pool is best compatible with an arrest in the synaptic-vesicle cycle after vesicle docking but before acquisition of fusion competence. A decrease in the total number of vesicles per terminal or in the docking process would have been detected in the morphometric analysis.

As neither calcium influx induced by action potentials, nor calcium ionophores or hypertonic sucrose solution are strong enough stimuli to induce release from Munc13-1-deficient synapses, we examined the effects of α -latrotoxin as a more drastic exocytotic stimulant. α -Latrotoxin, a component of black widow spider venom, is thought to induce transmitter release from nerve terminals by a dual mechanism involving both calcium influx through toxin-induced channels and a novel G-protein-coupled toxin receptor that ultimately stimulates phospholipase C^{13,14}. We found that, in Munc13-1-deficient neurons, spontaneous transmitter release induced by α -latrotoxin was similar to that of wild-type cells (Fig. 3e, f). This observation supports our morphological data, demonstrating that Munc13-1-deficient cultures create normal numbers of synapses. These contain synaptic vesicles loaded with transmitter that can exocytose upon α -latrotoxin treatment. Thus, Munc13-1 is not part of the fusion apparatus *per se*. Given that the Munc13-1-mutation disrupts synaptic vesicle maturation to fusion competence, we assume that α -latrotoxin itself stimulates or mediates a maturation process that compensates for or bypasses loss of Munc13-1. This effect of α -latrotoxin is not mediated by influx of calcium through toxin-induced channels, as calcium ionophores do not evoke a comparable effect in Munc13-1-deficient cells (Fig. 1d, e).

The maturation step in the synaptic vesicle cycle has been inferred from two key observations: (i) calcium-triggered exocytosis is too fast to be a complex multistep reaction; and (ii) during repetitive stimulation, exocytosis slows down before the number of docked vesicles declines^{1,2}. In adrenal chromaffin cells, biochemical assay paradigms have led to the identification of proteins involved in ATP-dependent maturation, or priming, of chromaffin granules¹⁵. In contrast, only sparse evidence exists to support the vesicle maturation postulate for central synapses, and proteins mediating synaptic vesicle maturation in the central nervous system are not known. Our electrophysiological and morphological observations of Munc13-1-deficient hippocampal neurons provide direct evidence for a maturation step in the synaptic-vesicle cycle of synapses in the central nervous system. The phenotype of mutant cells allows us to identify Munc13-1 as an essential protein for synaptic vesicle maturation. The phenotype of Munc13-1-deficient hippocampal neurons reveals an unexpected and striking heterogeneity between synapses with respect to synaptic-vesicle maturation. This heterogeneity is immediately obvious when normal inhibitory/GABA-mediated and strongly impaired excitatory/glutamatergic synapses are compared (Figs 1, 2). A similarly distinct heterogeneity with respect to Munc13-1-mediated vesicle maturation is apparent in the population of excitatory synapses of individual neurons. Although most synapses of any given single excitatory neuron (measured in microdot culture) are silent after loss of Munc13-1, transmitter release from a small subpopulation of synapses of the same cell is indistinguishable from wild-type controls and therefore independent of Munc13-1 (Fig. 3). These differences with respect to Munc13-1-mediated vesicle maturation are unlikely to be caused by simple compensatory mechanisms involving the other two existing Munc13-isoforms, Munc13-2 and Munc13-3, because GABA-mediated transmission in Munc13-2 and Munc13-3 deletion-mutant mice is not altered (C.R., I.A. and N.B., unpublished observations). Rather, the present data support the view that glutamatergic and GABA-mediated synapses utilize different vesicle-maturation and trafficking mechanisms. A similar general difference in maturation mechanisms between synapses may explain the fact that certain excitatory synapses depend on Munc13-1, but other synapses of the same cell do not. □

Methods

Munc13-1 deletion-mutant mice. Mutant mice were generated by homologous recombination in embryonic stem cells (see Supplementary Information). In the targeting vector, four exons, representing base pairs 2,587–2,964 of rat Munc13-1, were replaced with a neomycin resistance gene. Upon homologous recombination of the targeting vector with the Munc13-1 gene, the insertion of the neomycin resistance cassette results in the deletion of amino-acid residues 494–618 (including the C₁ domain) and a shift in the open reading frame of the mature messenger RNA. We identified recombinant stem cells and mice by Southern blotting and/or PCR. The absence of full-length and truncated Munc13-1 in homozygous mutant mice was confirmed by immunoblotting with antibodies directed against the extreme amino or C terminus of the protein. For all experiments, Munc13-1-deficient mice were obtained by interbreeding of animals heterozygous for the Munc13-1 mutation. Only wild-type or heterozygous littermates were used as controls.

Electrophysiology and cell culture. We prepared microdot cultures according to a modified version of published procedures^{6,9,11,16}. Experiments were performed using hippocampal neurons after 7–14 days in culture. To examine synaptically-activated currents, we used recurrent synapses (autapses) on microisland cultures. Only dots containing a single neuron were used. The extracellular medium contained (in mM): NaCl, 172; KCl, 2.4; HEPES, 10; glucose, 10; CaCl₂, 4; MgCl₂, 4; (350 mOsm, pH 7.3). Patch-pipette solutions included (in mM): KCl, 135; HEPES, 10; EGTA, 1; MgCl₂, 4.6; Na-ATP, 4; creatine phosphate, 15; phosphocreatine kinase, 50 U ml⁻¹; (325 mOsm, pH 7.3). Error bars indicate standard error. Statistical significance was tested using either Students *t*-test or one-way analysis of variance with the Bonferroni–Dunn procedure for multiple comparisons. Chemicals were from Sigma except for α -Latrotoxin (Alomone Labs), MK-801 (RBI), NMDA, GABA and kainate (Tocris Cookson).

Morphometry. To determine synapse densities, cultures were fixed as described⁴ and double stained with primary antibodies to synaptophysin (Synaptic Systems) and MAP-2 (Boehringer Mannheim) followed by the appropriate fluorescently labelled secondary antibodies. The number of synaptophysin-positive presynaptic terminals per μ m MAP-2-positive dendrite was counted in two independent wild-type and two independent Munc13-1-deficient cultures (total of 209 synapses in $-/-$ and 262 synapses in $+/+$ cultures). To determine the density of glutamatergic synapses, we double stained cultures with antibodies to GAD-65 (BioTrend) and synapsin (Synaptic Systems). GAD-65-negative/synapsin-positive varicosities were assumed to represent glutamatergic terminals (404 and 341 synapses in two $+/+$ and two $-/-$ cultures, respectively). Ultrastructural analyses were done as described¹⁷. Only asymmetric, that is, glutamatergic, synapses with clearly identifiable postsynaptic densities were analysed. We analysed the size of active zones, the number of docked vesicles (as defined by a direct contact between vesicle and plasma membrane) per μ m active zone length, and the number of vesicles within 150 nm of the active zone (expressed as number of vesicles per μ m active zone length) (122 synapses from 4 wild-type animals, 54 synapses from 2 knockout animals). Error bars indicate standard deviation.

Received 5 May; accepted 7 June 1999.

- Südhof, T. C. The synaptic vesicle cycle: a cascade of protein–protein interactions. *Nature* **375**, 645–653 (1995).
- Zucker, R. S. Exocytosis: a molecular and physiological perspective. *Neuron* **17**, 1049–1055 (1996).
- Brose, N., Hofmann, K., Hata, Y. & Südhof, T. C. Mammalian homologues of *C. elegans* unc-13 gene define novel family of C₂-domain proteins. *J. Biol. Chem.* **270**, 25273–25280 (1995).
- Betz, A. *et al.* Munc13-1 is a presynaptic porphol ester receptor that enhances neurotransmitter release. *Neuron* **21**, 123–136 (1998).
- Stevens, C. F. & Tsujimoto, T. Estimates for the pool size of releasable quanta at a single central synapse and for the time required to fill a pool. *Proc. Natl Acad. Sci. USA* **92**, 846–894 (1995).
- Rosenmund, C. & Stevens, C. F. Definition of the readily releasable pool of vesicles at hippocampal synapses. *Neuron* **16**, 1197–1207 (1996).
- Zucker, R. S. Changes in the statistics of transmitter release during facilitation. *J. Physiol.* **229**, 787–810 (1973).
- Zucker, R. S. Short-term synaptic plasticity. *Annu. Rev. Neurosci.* **12**, 13–31 (1989).
- Rosenmund, C., Clements, J. D. & Westbrook, G. L. Nonuniform probability of glutamate release at a hippocampal synapse. *Science* **262**, 754–757 (1993).
- Hessler, N. A., Shirke, A. M. & Malinow, R. The probability of transmitter release at a mammalian synapse. *Nature* **366**, 569–572 (1993).
- Rosenmund, C., Feltz, A. & Westbrook, G. L. Synaptic NMDA receptor channels have a low open probability. *J. Neurosci.* **15**, 2788–2795 (1995).
- Geppert, M. *et al.* Synaptotagmin I: a major Ca²⁺-sensor for transmitter release at a central synapse. *Cell* **79**, 717–727 (1994).
- Krasnoperov, V. G. *et al.* α -Latrotoxin stimulates exocytosis by the interaction with a neuronal

- G-protein-coupled receptor. *Neuron* **18**, 925–937 (1997).
14. Lelianaova, V. G. et al. α -Latrotoxin receptor, latrophilin, is a novel member of the secretin family of G-protein coupled receptors. *J. Biol. Chem.* **272**, 21504–21508 (1997).
 15. Martin, T. F. Mechanisms of protein secretion in endocrine and exocrine cells. *Vitamins Hormones* **54**, 207–226 (1998).
 16. Bekkers, J. M. & Stevens, C. F. Excitatory and inhibitory autaptic currents in isolated hippocampal neurons maintained in culture. *Proc. Natl Acad. Sci. USA* **88**, 7834–7838 (1991).
 17. Boyer, C., Schikorski, T. & Stevens, C. F. Comparison of hippocampal dendritic spines in culture and in brain. *J. Neurosci.* **18**, 5294–5300 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank J. Rettig and U. Ashery for stimulating discussions; S. Röglin, A. Kanbach and C. Casper for help with animal colonies; J. Finner and R. Schubert for artwork; I. Herfort, S. Schmidt and S. Wenger for technical assistance; and F. Benseler and I. Thanhäuser for DNA synthesis and sequencing. This work was supported by grants from the German Research Foundation and the Max-Planck-Society. C.R. and N.B. are Heisenberg Fellows of the German Research Foundation.

Correspondence and requests for materials should be addressed to N.B. (e-mail: brose@mail.mpiem.gwdg.de).

Cohesin Rec8 is required for reductional chromosome segregation at meiosis

Yoshinori Watanabe* & Paul Nurse

Cell Cycle Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Field, London WC2A 3PX, UK

When cells exit from mitotic cell division, their sister chromatids lose cohesion and separate to opposite poles of the dividing cell, resulting in equational chromosome segregation. In contrast, the reductional segregation of the first stage of meiotic cell division (meiosis I) requires that sister chromatids remain associated through their centromeres and move together to the same pole. Centromeric cohesion is lost as cells exit from meiosis II and sister chromatids can then separate^{1–4}. The fission yeast cohesin protein Rec8 is specific to and required for meiosis^{5–8}. Here we show that Rec8 appears in the centromeres and adjacent chromosome arms during the pre-meiotic S phase. Centromeric Rec8 persists throughout meiosis I and disappears at anaphase of meiosis II. When the *rec8* gene is deleted, sister chromatids separate at meiosis I, resulting in equational rather than reductional chromosome segregation. We propose that the persistence of Rec8 at centromeres during meiosis I maintains sister-chromatid cohesion, and that its presence in the centromere-adjacent regions orients the kinetochores so that sister chromatids move to the same pole. This results in the reductional pattern of chromosome segregation necessary to reduce a diploid zygote to haploid gametes.

Mutations in the *rec8* gene of the fission yeast *Schizosaccharomyces pombe* cause decreased meiotic recombination in centromeric regions^{5,9}, precocious segregation of sister chromatids at meiosis I and disruption of linear elements (structures implicated in chromosome pairing)⁶. It has been suggested from gene-sequence comparisons that Rec8 is a member of the cohesin family of proteins⁷. We have confirmed this by cloning and sequencing a *rec8* complementary DNA, and by showing that it potentially encodes a protein of 561 amino acids with significant homology to cohesins, including the fission-yeast mitotic cohesin Rad21 (Fig. 1a). The carboxy terminus of endogenous *rec8* was tagged with either haemagglutinin (HA) or green fluorescent protein (GFP), and both tagged Rec8 constructs were shown to be fully functional by their ability to rescue the meiotic defects of a *rec8* mutant (data not shown).

Rec8–HA was absent in mitotically growing cells, but was present during meiosis, appearing just before the pre-meiotic S phase (Fig. 1b). Similar results were found with Rec8–GFP (Fig. 2a).

The protein undergoes phosphorylation as cells proceed through the later stages of meiosis (Fig. 1b). To determine whether Rec8 has cohesin-like properties, we tested whether ectopically expressing Rec8–GFP could rescue the mitotic growth of cells deleted for the mitotic cohesin gene *rad21*. Moderate expression of Rec8–GFP allowed cells deleted for *rad21* to grow (Fig. 1c), although growth rate was reduced by 20% and some inviable cells were formed. We conclude that Rec8 has basal cohesin activity and is a cohesin that is present only during meiosis. In contrast, overexpression of Rad21–GFP could not rescue the meiotic defects of *rec8Δ* (data not shown), indicating that Rec8 may also have specialized activities required specifically for meiosis (see below).

To investigate the role of Rec8 in meiotic chromosome segregation, we deleted the *rec8* gene and monitored the segregation of a tagged chromosome containing a LacO DNA repeat integrated at the *lys1* locus, 4 cM from the centromere of chromosome I (ref. 10). Segregation of this centromeric region was monitored microscopically by expressing a LacI–GFP fusion to target the LacO repeat. Crosses were made between one strain with the integrated LacO tag and another strain without the tag. A cross between *rec8*⁺ and *rec8Δ* undergoes normal meiosis and was used as a control. In this cross, the tagged sister centromeres mostly moved to the same pole during meiosis I, and separated during meiosis II (Fig. 3a). In contrast, in a cross between *rec8Δ* and *rec8Δ*, the tagged sister centromeres almost all separated during meiosis I (Fig. 3a). This indicates that, in the absence of the Rec8 cohesin, cells undergo equational chromosome segregation at meiosis I.

To investigate this possibility further, we used a *cdc2*^{ts} mutant to block meiosis II. In this mutant, two spored dyads are formed as a consequence of meiosis I being completed and meiosis II being blocked¹¹. To assess whether chromosome segregation at meiosis I was reductional, we monitored the segregation of three centromere markers, *lys1*, *pat 1* (5 cM from chromosome II)¹² and *ade6* (16 cM from chromosome III centromere)¹³. In *rec8*⁺ cells, the dyads were at least 80% viable with predominantly +/+, +/– genotypes for all three centromere-linked markers, indicative of reductional segregation (Fig. 3b, c). The occurrence of some +/–, +/– dyads can be explained by recombination between the markers and the centromeres. In *rec8Δ* cells, the dyads generated were also at least 80% viable but had exclusively +/–, +/– genotypes for all three markers, indicative of equational segregation (Fig. 3b, c). When tetrad formation was monitored in the *rec8Δ* cells, spore viability dropped to 21% (Fig. 3b), and 4,6-diamidino-2-phenylindole (DAPI) staining of nuclei showed that chromosome segregation was equal at

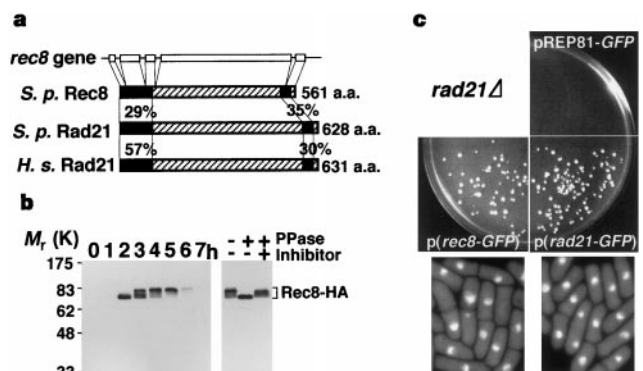


Figure 1 The *rec8* gene encodes a meiosis-specific cohesin and is phosphorylated. **a**, *rec8* consists of five exons. *S. pombe* (*S.p.*) mitotic cohesin Rad21 (ref. 28) is aligned with *S. pombe* Rec8 and a human (*H.s.*) Rad21 homologue²⁹. Black bars indicate the regions of highest shared similarity. The degree (%) of identity between *S.p.* Rad21 and *S.p.* Rec8 or *H.s.* Rad21 is shown. **b**, Rec8 is transiently expressed during meiosis and is phosphorylated. **c**, Haploid *rad21Δ* carrying *p(rec8-GFP)* as well as *p(rad21-GFP)* formed colonies but those carrying *pREP81-GFP* did not. Proliferating transformants were stained by DAPI.

* Present address: Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Hongo, Tokyo 113, Japan.

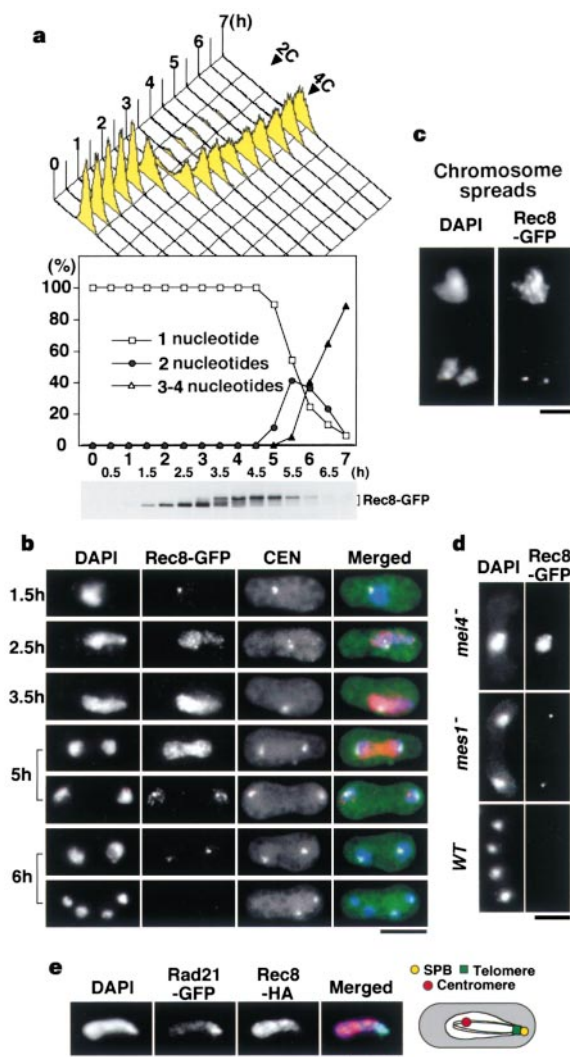


Figure 2 Rec8 associated with centromeres and with centromere-proximal chromosome arm regions during meiosis. **a**, Synchronous meioses of diploid *rec8-GFP* cells were sampled every 30 min and DNA content measured by FACSscan (top), meiotic nuclear division monitored by DAPI staining (middle) and protein level and phosphorylation of Rec8 detected by western blotting using anti-GFP antibodies (bottom). **b**, The same samples were examined by FISH using the centromere probe, anti-GFP antibodies and DAPI. Representative cells at each time point are shown. In the merged figure, DAPI, Rec8-GFP and centromeres are represented by blue, red and green, respectively. Note that two centromeres are out of focus in the 4-nuclei cells at 6 h. **c**, Chromosome spreads were examined by indirect immunofluorescence for the localization of Rec8. Cells at metaphase I (top) and anaphase I (bottom) are shown. **d**, *h⁹⁰ rec8-GFP* cells of *mei4Δ*, *mei1-B44* and wild type (WT) were sporulated and examined for DAPI and GFP signals. **e**, Localization of Rad21-GFP and Rec8-HA in a meiotic horse-tail nucleus. In the merged figure, DAPI, Rad21-GFP and Rec8-HA are represented by blue, green and red, respectively. The positions of SPB, telomere and centromere are shown in the drawing of the horse-tail nucleus. Bars indicate 5 μ m.

meiosis I but unequal at meiosis II (Fig. 3d). We conclude that, when the meiosis-specific cohesin Rec8 is absent, the centromere cohesion of sister chromatids does not persist throughout the meiosis I and so the reductional segregation pattern is completely shifted to an equational segregation pattern. The lack of cohesion during the subsequent meiosis II leads to random chromosome segregation.

If Rec 8 is important for cohesion of sister chromatid centromeres, then it might be expected to be located in the centromeric regions of chromosomes. We tested this using Rec8-GFP. Before pre-meiotic S phase, Rec8-GFP appeared as a single dot in the nucleus, and FISH

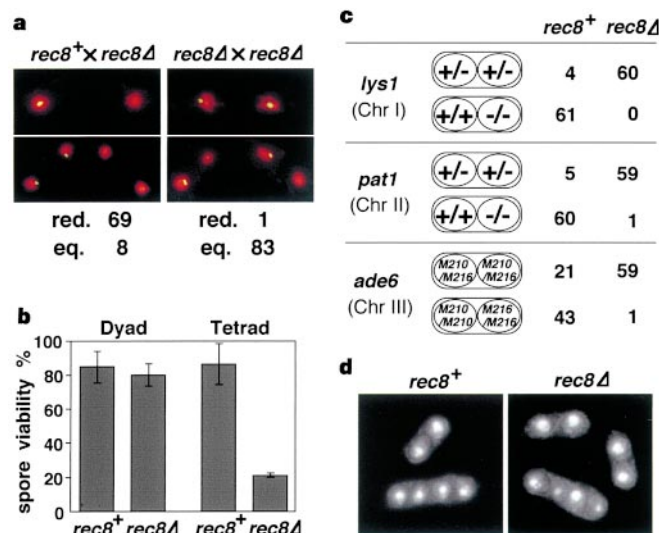


Figure 3 Shift in behaviour from reductional to equational chromosome segregation during meiosis I in *rec8Δ*. **a**, Segregation of GFP fluorescence associated with the *lys1* locus during meiosis I and meiosis II. Green and red represent GFP and DAPI, respectively. The number of cells in which segregation pattern of GFP was reductional (red.) or equational (eq.) is given. **b-d**, Diploid *rec8⁺* and *rec8Δ* strains carrying a *cdc-2-L7* allele were sporulated on YPD plates. Dyads and tetrads were dissected and examined for spore viability (**b**). Fully viable dyads were analysed for the segregation of the centromere-linked markers (**c**). Dyads and tetrads were stained by DAPI (**d**).

was used to show that this dot was located with the centromeres that cluster together at this point in meiosis (Fig. 2a, b; 1.5 h). At prophase of meiosis I, the nucleus takes on a horse-tail shape and moves backwards and forwards¹⁴. At this time, Rec8-GFP extends further through the nucleus (Fig. 2b; 2.5 h), and chromosome spreads show that it is tightly associated with chromatin (Fig. 2c, upper cell). A similar localization was observed in a *mei4Δ* mutant that blocks at the same stage of meiosis¹⁵ (Fig. 2d). During the horse-tail movements the telomeres are clustered at the leading edge of the moving elongated nucleus, whereas the centromeres are found at the other end of the nucleus¹⁴ (Fig. 2e). Rec8-GFP was less prevalent at the leading edge of the nucleus, where the telomeres are clustered, and was more concentrated in the regions of the nucleus containing the centromeres (Fig. 2b, e). As cells proceeded through meiosis I into meiosis II, the level of Rec8-GFP detected by western blotting was reduced (Fig. 2a), and much of the Rec8-GFP became dissociated from chromatin regions (Fig. 2b; 5 h). However, Rec8-GFP remained tightly associated with centromeric chromatin regions until metaphase of meiosis II (Fig. 2b; 5 h, 6 h; Fig. 2c, bottom cell), and can be detected as a single dot in both nuclei of cells blocked at meiosis II using a *mes1* mutant¹⁶ (Fig. 2d). Substantial Rec8-GFP occasionally remains in the nucleus at anaphase of meiosis I (Fig. 2b; 5 h, upper cell); however, nuclear spreads of dividing chromosome detected no associated Rec8-GFP except at the centromeres (Fig. 2c, bottom cell). These results fit a model in which the dissociation of Rec8 from the arms could trigger the onset of anaphase during meiosis I (Fig. 4). By anaphase of meiosis II, no Rec8-GFP could be detected either by western blotting (Fig. 2a) or cytologically (Fig. 2b; 6 h). Similar behaviour was found using Rec8-HA (Fig. 2e, and data not shown). These observations indicate that Rec8 is located in the regions of the centromeres, with a more general distribution throughout the nucleus during meiosis I. It disappears completely by the end of meiosis II.

To examine the distribution of the mitotic cohesin Rad21 during meiosis, we used a *rec8-HA* strain carrying a plasmid expressing Rad21-GFP. In mitotically dividing cells, Rad21-GFP was distributed throughout the entire nucleus (data not shown) as previously

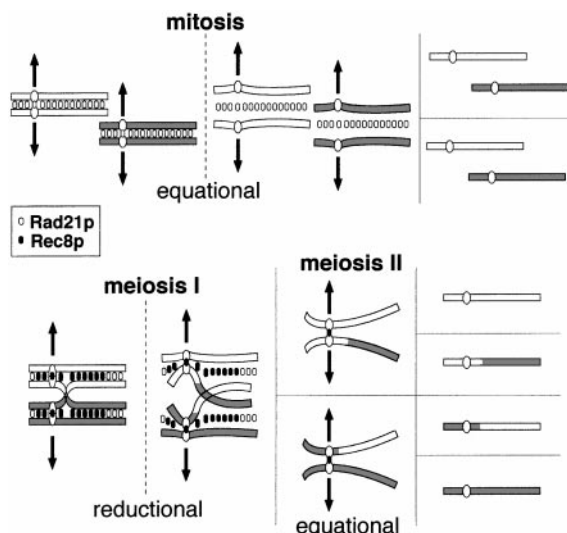


Figure 4 Model for the action of mitotic and meiotic cohesins. The behaviour of homologous chromosomes (white and black) and cohesins (Rad21 white; Rec8, black) in mitosis and meiosis is depicted. The large oval on the chromosome indicates the centromere. At anaphase of mitosis, sister-chromatid cohesion is released in both arm and centromeric regions at once. In meiosis, cohesion along the chromosome arm is lost during meiosis I to release recombined homologue connections, whereas centromeric cohesion is maintained until meiosis II. Note that kinetochores of sister pairs are modified to face the same pole in meiosis I, although facing the opposite poles in mitosis and meiosis II.

reported¹⁷, but during meiosis, Rad21-GFP was found predominantly at the leading edge of the meiosis I prophase nuclei, where the telomeres are clustered (Fig. 2e). This contrasts with the global distribution of Rec8-HA around the centromeric regions. Thus during meiosis I, the mitotic cohesin Rad21 is located mainly towards the telomeric regions rather than the centromeric regions, whereas the meiotic cohesin Rec8 is located mainly in the centromeric regions rather than the telomeric regions. This polarized localization of Rec8 may explain why pairing and recombination are impaired in the centromere-adjacent regions but not in the telomere-adjacent regions in a *rec8* mutant^{5,6,9}.

These results show that Rec8 is required to shift chromosome segregation from being equational during the mitotic cell cycle to being reductional during the meiotic cell cycle. We therefore next tested whether the expression of Rec8 during the mitotic cell cycle might influence the pattern of chromosome segregation. We monitored the segregation of the *ade6* chromosomal centromere-III marker in diploid *rad21Δ* cells expressing moderate levels of either the mitotic cohesin Rad21-GFP as a control, or Rec8-GFP. After culturing for six generations with adenine, 7% of $-/-$ segregants (indicative of reductional segregation) were observed in the Rec8-GFP-expressing cells, compared with 0.1% in the control Rad21-GFP-expressing cells. This result indicates that when Rec8 replaces Rad21 during the mitotic cell cycle, the chromosome segregation pattern is shifted in a significant number of cells from being equational to being reductional. This is consistent with the observation that haploid *rad21Δ* cells carrying *p(rec8-GFP)* frequently undergo unequal chromosome segregation, producing inviable cells (Fig. 1c).

During the mitotic cell cycle, Rad21 in fission yeast and its homologue Scc1 in budding yeast associate with the whole length of sister chromatids and establish sister chromatid cohesion^{7,17,18}. This happens during S phase and is likely to be coupled with DNA replication¹⁸⁻²⁰. At mitotic anaphase, Scc1 dissociates from chromatin, allowing the chromatids to separate and segregate to opposite poles of the dividing cell, resulting in equational segregation²¹. Rad21 may work similarly during the mitotic cell

cycle of fission yeast (Fig. 4). Based on our results, we propose the following model of meiotic chromosome segregation. At premeiotic S phase, Rec8 is incorporated into the centromeres and centromeric regions of the sister chromatid arms. This may occur as the centromere regions are replicated during pre-meiotic S phase. Changing the chromosome segregation pattern from equational to reductional requires the sister-chromatid centromeres to remain associated throughout meiosis I so that the sister chromatids segregate to the same pole, and for the centromeres to become dissociated at anaphase of meiosis II to allow sister chromatids to separate to opposite poles (Fig. 4). The persistence of the Rec8 cohesin at centromeres throughout meiosis I would maintain sister-chromatid-centromere cohesion, and its disappearance at the end of meiosis II would allow sister-chromatid-centromere separation (Fig. 4). The shift from equational to reductional segregation also requires a change in orientation of the sister-chromatid kinetochores built on the centromeres. For equational segregation to occur, the kinetochores must be oriented towards opposite poles, whereas for reductional segregation they must be oriented towards the same pole². Given that Rec8 is required for the shift to reductional segregation at meiosis I, we propose that the presence of Rec8 in the chromosome arms adjacent to the centromeres coordinates the orientation of sister-chromatid kinetochores so that they are directed towards the same pole. The dissociation of Rec8 from the chromosome arms at meiosis II would then allow the kinetochores of sister chromatids to become oriented towards opposite poles, resulting in the chromatids moving apart (Fig. 4).

In the absence of Rec8, faithful equational chromosome segregation is still observed, indicating that sister-chromatid cohesion must be maintained until metaphase of meiosis I. Consistent with this conclusion, fluorescent *in situ* hybridization (FISH) experiments in a *rec8* mutant indicated that the clustering and cohesion of centromeres were intact during meiotic prophase, whereas the pairing and cohesion of the centromere-adjacent chromatid arms appeared to be impaired⁶. It is not known what maintains this centromeric cohesion, but it could involve other non-cohesin proteins such as MeiS-322, which may be required for centromere cohesion during *Drosophila* meiosis²².

Here we have shown that the meiotic cohesin Rec8 is important for establishing reductional kinetochore behaviour at meiosis. A putative Rec8 homologue has been cloned from human cells⁹, indicating that the mechanism for establishing reductional segregation proposed here may be relevant for meiosis in metazoan organisms. Problems during meiosis that result in precocious sister-chromatid separation are important for certain human diseases, including Down's syndrome and infertility, indicating that the involvement of meiosis-specific cohesins in these diseases should be investigated. □

Methods

Media and strains. All media and growth conditions are as described²³ unless otherwise stated. For sporulation agar medium we used YPD (1% yeast extract, 2% polypeptone and 2% dextrose) or EMM (glutamate), which contains sodium glutamate (0.5 g l⁻¹) in place of NH₄Cl. The adenine auxotrophic mutations *ade6-M210* and *ade6-M216* can complement each other; homozygous *ade6-M210/ade6-M210* and *ade6-M216/ade6-M216* diploids form pink colonies on adenine-limiting (10 μg l⁻¹) plates, and heterozygous *ade6-M210/ade6-M216* diploids form white colonies.

Complementation of *rad21Δ* deletion. Diploid *h⁺/h⁻ rad21::ura4⁺/rad21⁺ ura4-D6/ura4-D6 leu/leu1 ade6-M210/ade6-M216* cells were transformed with *pREP81-GFP*, *p(rad21-GFP)* or *p(rec8-GFP)* and subjected to sporulation. After being treated with Helix Pomatia extract to kill mitotic diploid cells, spores were spread on EMM supplemented with adenine, where only *rad21Δ*-deleted cells (Ura⁺) carrying the plasmid (Leu⁺) could form colonies.

Gene disruption, tagging and plasmids. Deletion and HA or GFP tagging of endogenous *rec8* were performed using the PCR-based gene-targeting method for *S. pombe*²⁴. Both *rec8* and *rad21* with their promoter regions were amplified by PCR and cloned into *pREP81-GFP*, fusing GFP to their C-terminal ends in

frame to generate the plasmids p(*rec8*–GFP) and p(*rad21*–GFP).

Monitoring segregation of centromeres tagged with GFP during meiosis. *rec8Δ* cells tagged with GFP at the *lys1* locus were crossed with non-tagged *rec8⁺* cells or *rec8Δ* cells. For this purpose, both types of cell were cultured to log phase, collected by centrifugation, mixed and then spotted on a sporulation plate SPA. This allows subsequent synchronous conjugation and meiosis. When cells undergoing meiosis I and II became abundant in the population, cells were fixed with cold methanol and zygotes monitored for GFP and DAPI.

Spore analysis. *h⁺/h⁻ cdc2-L7 rec8Δ* diploid cells carrying heterozygous centromere-linked markers (*lys1*, *pat1* and *ade6*) and isogenic *rec8⁺* cells were sporulated on a YPD plate. At 33 °C they preferentially formed two spores that had undergone only meiosis I; at 25 °C they further underwent meiosis II and formed four spores. Dyads or tetrads were dissected and incubated on YE plates at 25 °C for 7 days. Fully viable dyad pairs were further tested for auxotrophy by replica plating on minimal plates at 25 °C, and for the *pat1* phenotype by incubating the replicated plate at 33 °C.

Synchronous meiosis. To induce synchronous meiosis, we used a temperature sensitive *pat1-114* allele¹². For FISH experiments, a diploid strain *h⁺/h⁺ cdc2-L7/ cdc2-L7 pat1-114/pat1-114 rec8-GFP/rec8-GFP lys1-131/lys1⁺ ade6-M210/ade6-M216* was cultured in EMM-N at 25 °C for 16 h, and G1 cells were collected by elutriation. We placed these cells were at 34 °C and added 0.5 g l⁻¹ of NH₄Cl; after 2 h cultures were shifted to 30 °C.

Detection of Rec8 phosphorylation. Synchronous meiosis was induced in diploid *pat1 ced2 rec8–HA* cells. Cell extracts were prepared every 1 h and Rec8–HA was detected by western blotting using anti-HA antibodies (12CA5). Extracts from the 4-h sample were incubated with and without calf intestine alkaline phosphatase (Sigma), or with phosphatase plus a phosphatase inhibitor mix (60 mM β-glycerophosphate, 15 mM nitrophosphate, 1 mM NaV₃), and were analysed by western blotting.

GFP fluorescence and FISH. For GFP fluorescence, cells were fixed by cold methanol, stored at –20 °C, washed and suspended in PEMS buffer (100 mM PIPES, pH 6.9, 1 mM EGTA, 1 mM MgSO₄, 1.2 M sorbitol) with DAPI. The method for FISH in fission yeast was previously described²⁵. To digest the meiotic cell wall, we used ten times the concentration of the enzymes (zymolyase at 10 mg ml⁻¹ and novozyme at 2 mg ml⁻¹) and incubated at 36 °C for 30 min. We used a digoxigenin-labelled centromere repeat probe¹³ and purified anti-GFP antibodies (a gift from K. Sawin).

Chromosome spreads. Spread nuclei were prepared as described²⁶. For immunostaining, previous methods were followed²⁷, using anti-GFP antibodies diluted 1:50 in PBS for the first antibody, and Cy3-conjugated goat anti-rabbit IgG diluted 1:200 for the second antibody.

Assay of reductional segregation in mitotic cells expressing Rec8–GFP. Diploid *h⁺/h⁻ rad21::ura4⁺/rad21::ura4⁺ ura4-D6/ura4-D6 leu/leu1 ade6-M210/ade6-M216* cells carrying p(*rad21*–GFP) or p(*rec8*–GFP) cultured without adenine were moved to medium with adenine. After culturing for six generations, they were spread on adenine-limiting plates and examined for adenine auxotrophy. Cells that had undergone reductional segregation of the *ade6* allele could be detected because they formed pink colonies. By microscopic inspection of each pink colony, we excluded haploid or aneuploid colonies caused by chromosome loss or missegregation.

Received 31 March; accepted 8 June 1999.

- Miyazaki, W. Y. & Orr-Weaver, T. L. Sister-chromatid cohesion in mitosis and meiosis. *Annu. Rev. Genet.* **28**, 167–168 (1994).
- Bickel, S. E. & Orr-Weaver, T. L. Holding chromatids together to ensure they go their separate ways. *BioEssays* **18**, 293–300 (1996).
- Kleckner, N. Meiosis: how could it work? *Proc. Natl. Acad. Sci. USA* **93**, 8167–8174 (1996).
- Roeder, G. S. Meiotic chromosomes: it takes two to tango. *Genes Dev.* **11**, 2600–2621 (1997).
- DeVaux, L. C. & Smith, G. R. Region-specific activators of meiotic recombination in *Schizosaccharomyces pombe*. *Genes Dev.* **8**, 203–210 (1994).
- Molnar, M., Bahler, J., Sipiczki, M. & Kohli, J. The *rec8* gene of *Schizosaccharomyces pombe* is involved in linear element formation, chromosome pairing and sister-chromatid cohesion during meiosis. *Genetics* **141**, 61–73 (1995).
- Michaelis, C., Ciosk, R. & Nasmyth, K. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* **91**, 35–45 (1997).
- Guacci, V., Koshland, D. & Strunnikov, A. A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in *S. cerevisiae*. *Cell* **91**, 47–57 (1997).
- Parisi, S. *et al.* Rec8p, a meiotic recombination and sister chromatid cohesion phosphoprotein of the Rad21p family, conserved from fission yeast to humans. *Mol. Cell. Biol.* **19**, 3515–3528 (1999).
- Nabeshima, K. *et al.* Dynamics of centromeres during metaphase–anaphase transition in fission yeast: dis1 is implicated in force balance in metaphase bipolar spindle. *Mol. Biol. Cell* **9**, 3211–3225 (1998).
- Nakaseko, Y., Niwa, O. & Yanagida, M. A meiotic mutant of the fission yeast *Schizosaccharomyces pombe* that produces mature asci containing two diploid spores. *J. Bacteriol.* **157**, 334–336 (1984).

- Iino, Y. & Yamamoto, M. Mutants of *Schizosaccharomyces pombe* which sporulate in the haploid state. *Mol. Gen. Genet.* **198**, 416–421 (1985).
- Chikashige, Y. *et al.* Composite motifs and repeat symmetry in *S. pombe* centromeres. *Cell* **57**, 739–751 (1989).
- Chikashige, Y. *et al.* Telomere-led premeiotic chromosome movement in fission yeast. *Science* **264**, 270–273 (1994).
- Horie, S. *et al.* The *Schizosaccharomyces pombe meid⁴* gene encodes a meiosis-specific transcription factor containing a forkhead DNA-binding domain. *Mol. Cell. Biol.* **18**, 2118–2129 (1998).
- Kishida, M., Nagai, T., Nakaseko, Y. & Shimoda, C. Meiosis-dependent mRNA splicing of the fission yeast *Schizosaccharomyces pombe mes1⁺* gene. *Curr. Genet.* **25**, 497–503 (1994).
- Birkenbihl, R. P. & Subramani, S. The *rad21* gene product of *Schizosaccharomyces pombe* is a nuclear, cell cycle-regulated phosphoprotein. *J. Biol. Chem.* **270**, 7703–7711 (1995).
- Toth, A. *et al.* Yeast Cohesin complex requires a conserved protein, Eco1p(Ctf7), to establish cohesion between sister chromatids during DNA replication. *Genes Dev.* **13**, 320–333 (1999).
- Uhlmann, F. & Nasmyth, K. Cohesion between sister chromatids must be established during DNA replication. *Curr. Biol.* **8**, 1095–1101 (1998).
- Skibbens, R. V., Corson, L. B., Koshland, D. & Hieter, P. Ctf7p is essential for sister chromatid cohesion and links mitotic chromosome structure to the DNA replication machinery. *Genes Dev.* **13**, 307–319 (1999).
- Ciosk, R. *et al.* An ESP1/PDS1 complex regulates loss of sister chromatid cohesion at the metaphase to anaphase transition in yeast. *Cell* **93**, 1067–1076 (1998).
- Tang, T. T., Bickel, S. E., Young, L. M. & Orr-Weaver, T. L. Maintenance of sister-chromatid cohesion at the centromere by the *Drosophila* MEI-S322 protein. *Genes Dev.* **12**, 3843–3856 (1998).
- Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
- Bähler, J. *et al.* Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14**, 943–951 (1998).
- Cooper, J. P., Watanabe, Y. & Nurse, P. Fission yeast Taz1 protein is required for meiotic telomere clustering and recombination. *Nature* **392**, 828–831 (1998).
- Bähler, J., Wyler, T., Loidl, J. & Kohli, J. Unusual nuclear structures in meiosis prophase of fission yeast: a cytological analysis. *J. Cell Biol.* **121**, 241–256 (1993).
- Klein, F. *et al.* Localization of RAP1 and topoisomerase II in nuclei and meiotic chromosomes of yeast. *J. Cell Biol.* **117**, 9335–948 (1992).
- Birkenbihl, R. P. & Subramani, S. Cloning and characterization of *rad21* an essential gene of *Schizosaccharomyces pombe* involved in DNA double-strand-break repair. *Nucleic Acids Res.* **20**, 6605–6611 (1992).
- Nomura, N. *et al.* Prediction of the coding sequences of unidentified human genes. *DNA Res.* **1**, 223–229 (1994).

Acknowledgements. We thank M. Yamamoto, M. Yanagida and S. Subramani for plasmids and strains; J. Bähler, J. Cooper and J. Hayles for advice with the manuscript; and J. Kohli and M. McKay for communicating results before publication. Y.W. thanks all the members of P. Nurse's lab for advice and reagents, and the ICRF staff and M. Yamamoto for support. Y.W. was supported by JSPS and a Uehara fellowship.

Correspondence and requests for materials should be addressed to Y.W. (e-mail: ywatanab@ims.u-tokyo.ac.jp). The GenBank accession number of *rec8* cDNA sequence is ABO 18077.

Creation of human tumour cells with defined genetic elements

William C. Hahn*†‡, Christopher M. Counter†‡§, Ante S. Lundberg*†‡, Roderick L. Beijersbergen*, Mary W. Brooks* & Robert A. Weinberg*

* Whitehead Institute for Biomedical Research, 9 Cambridge Center, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

† Department of Adult Oncology, Dana-Farber Cancer Institute, Departments of Medicine, Brigham and Women's Hospital and Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA

‡ Department of Pharmacology and Cancer Biology, Department of Radiation Oncology, Box 3813, Duke University Medical Center, Durham, North Carolina 27710, USA

During malignant transformation, cancer cells acquire genetic mutations that override the normal mechanisms controlling cellular proliferation. Primary rodent cells are efficiently converted into tumorigenic cells by the coexpression of cooperating oncogenes^{1,2}. However, similar experiments with human cells have consistently failed to yield tumorigenic transformants^{3–5}, indicating a fundamental difference in the biology of human and rodent cells. The few reported successes in the creation of human tumour cells have depended on the use of chemical or physical agents to achieve immortalization⁶, the selection of rare, spontaneously arising immortalized cells^{7–10}, or the use of an entire viral genome¹¹. We show here that the ectopic expression of the

§ These authors contributed equally to this work.

telomerase catalytic subunit (*hTERT*)¹² in combination with two oncogenes (the simian virus 40 large-T oncoprotein and an oncogenic allele of *H-ras*) results in direct tumorigenic conversion of normal human epithelial and fibroblast cells. These results demonstrate that disruption of the intracellular pathways regulated by large-T, oncogenic *ras* and telomerase suffices to create a human tumor cell.

One ostensibly important difference between rodent and human cells derives from their telomere biology. Murine somatic cells express telomerase activity and have much longer telomeres¹³ than their normal human counterparts, which lack telomerase activity^{14,15}. Because normal human cells progressively lose telomeric DNA with passage in culture, telomeric erosion is thought to limit cellular lifespan¹⁶. Ectopic expression of the *hTERT* gene, which encodes the catalytic subunit of the telomerase holoenzyme, enables some¹⁷ but not all (ref. 18, W. C. Hahn and R. A. Weinberg, unpublished observations) pre-senescent primary human cells to multiply indefinitely. Moreover, ectopic *hTERT* enzyme expression enables telomerase-negative, transformed cells to bypass crisis, a proliferative barrier after senescence characterized by widespread cell death¹⁹. Together, these studies indicate that telomerase can confer replicative immortality on certain human cell types. Because almost all human tumours have detectable telomerase activity¹⁵, we reasoned that telomerase might contribute to the tumorigenic

conversion of human cells.

To determine whether human cells immortalized by the ectopic expression of *hTERT* were tumorigenic, we serially introduced combinations of *hTERT* and an oncogenic *ras* (*H-rasV12*) allele²⁰ by using amphotropic retroviruses into human embryonic kidney (HEK) cells that express large-T to bypass senescence (HA1 cells)²¹ (Fig. 1). We also introduced large-T and combinations of *hTERT* and oncogenic *ras* into early-passage (passage 5) normal human BJ fibroblasts (Fig. 1). For each infection, parallel cultures were infected with control retroviruses specifying only a drug resistance gene as controls. Polyclonal, mass-infected populations as well as clonal isolates were obtained for each combination of serially introduced genes. We estimate that approximately 60 population doublings were expended during the process of serially introducing large-T, *hTERT* and *ras*. We could not propagate cells expressing only *hTERT* and *ras*, as these cells entered a senescent state immediately after the introduction of *ras*²⁰, confirming that *hTERT* does not abrogate a *ras*-induced growth arrest (data not shown)²². Amounts of large-T expressed were similar in all cell populations (Fig. 1a) and cell clones (data not shown), and increased expression of the Ras protein was observed in the expected cell populations (Fig. 1b) and clones (data not shown) after infection with a vector transducing the *ras* oncogene. We also confirmed that the ectopic expression of *hTERT* (Fig. 1c) resulted in telomerase activity in both HEK and BJ cells (Fig. 1d). This telomerase activity resulted in both telomere elongation (Fig. 2a) and stabilization (Fig. 2b) in these cells, as assessed by Southern blotting of genomic DNA. Expression of the *ras* oncogene did not affect the ability of telomerase to maintain telomeres in these cells (Fig. 2a, b).

Expression of *hTERT* led to the immortalization of BJ fibroblasts expressing large-T, recapitulating the results seen previously with HA1 and pre-senescent BJ cells (Fig. 3)^{17,21}. In contrast, telomerase-negative polyclonal (Fig. 3a, b) and monoclonal (data not shown) cell populations expressing only large-T and *ras* entered crisis within

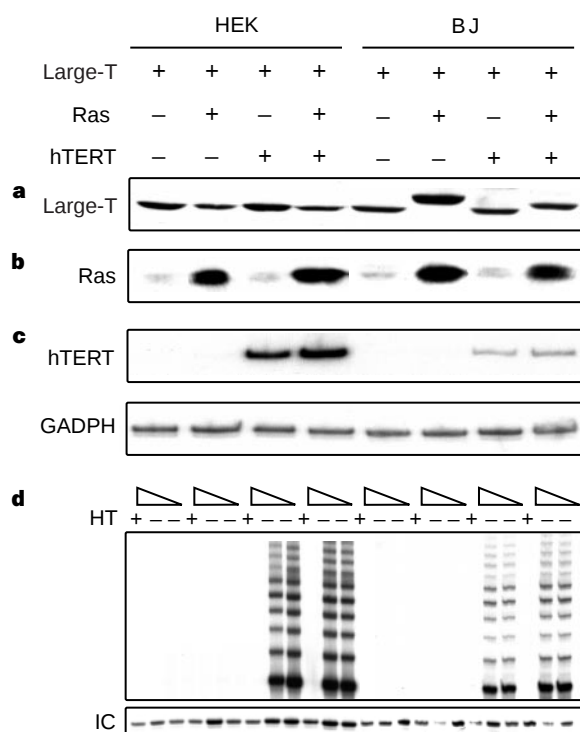


Figure 1 Expression of large-T, *ras* and telomerase. Immunoblotting for the large-T of relative molecular mass 80,000 (M_r 80K) (**a**) and M_r 21K Ras (**b**) proteins confirmed equal expression of large-T and overexpression of Ras in the indicated cells. The anti-Ras antibody (259) used in this study recognizes a determinant common to both wild-type and mutated Ras. Large-T was not detected in uninfected HEK and BJ cells (data not shown). The slower migration of large-T in the presence of *ras* expression was seen only in BJ cells and was observed with two different anti-large-T monoclonal antibodies. **c**, Expression of *hTERT* transcripts was confirmed by RT-PCR. The *hTERT*-specific primers amplified a 175 bp product. Primers specific for glyceraldehyde-3-phosphate dehydrogenase (GADPH) (220 bp product) confirmed that equal amounts of RNA were present in each sample. **d**, Cellular extracts (200 and 20 ng) were tested for telomerase activity by using the PCR-based TRAP assay; heat-treated samples (HT) were included as controls. IC refers to an internal PCR standard to demonstrate the absence of PCR inhibitors in the cellular extracts.

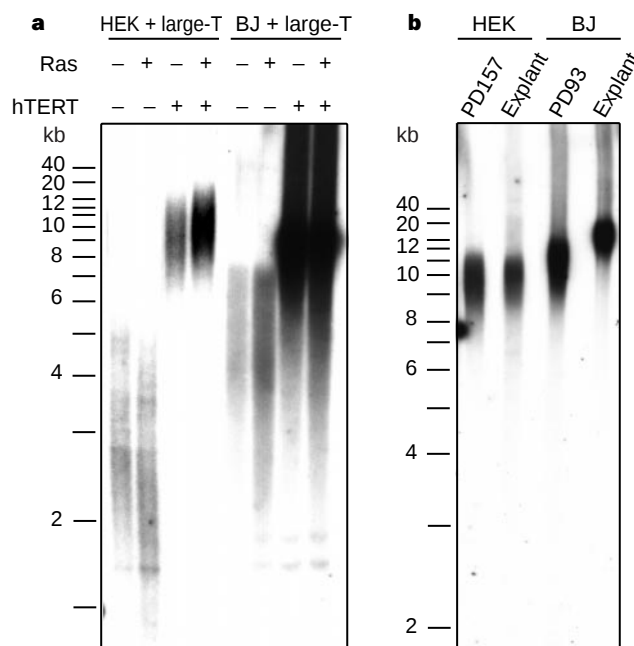


Figure 2 Expression of *hTERT* stabilizes telomere length. **a**, Telomere length for HEK and BJ cells was analysed by hybridization of genomic DNA with a telomere-specific oligonucleotide probe after 6 population doublings. The positions of size standards (kb) are indicated at the left. **b**, Telomere lengths for HEK and BJ cells expressing large-T, *hTERT* and *ras* at late passage (HEK, population doubling 157; BJ population doubling 93) and after recovery (explant) from a tumour formed in a nude mouse.

Table 1 Anchorage-independent growth of large-T-expressing HEK and BJ cells

Experiment 1*									
No. of cells seeded	Number of colonies after 21 d								Calu-1 (control)
	HEK + LT				BJ + LT				
	hTERT ⁻ Ras ⁻	hTERT ⁻ Ras ⁺	hTERT ⁺ Ras ⁻	hTERT ⁺ Ras ⁺	hTERT ⁻ Ras ⁻	hTERT ⁻ Ras ⁺	hTERT ⁺ Ras ⁻	hTERT ⁺ Ras ⁺	
	10 ⁴	10	40	4	160	0	3	0	
10 ³	0	12	0	80	0	0	0	40	200 75
Experiment 2†									
No. of cells seeded	Number of colonies after 21 d					293 (control)			
	HEK; LT + hTERT + Ras		BJ; LT + hTERT + Ras						
	Parental	Explanted	Parental	Explanted					
	10 ⁵	3,460	2,850	4,046	4,858				
10 ⁴	785	434	524	492	2,220 286				
10 ³	106	34	50	64	29				

The lung cancer cell line Calu-1 and the adenovirus-transformed cell line 293 were used as positive controls.

* Anchorage-independent growth of HEK and BJ cells with combinations of large-T (LT), *hTERT*, and *ras*. Six clonal isolates from HEK or BJ cells expressing only LT and *hTERT* never formed colonies, whereas clonal isolates from HEK or BJ cells expressing LT, *hTERT* and *ras* produced variable numbers of colonies (HEK 0–40 and BJ 7–247; 10⁴ cells seeded). The number of colonies formed by such cells correlated with the expression of *ras* in each clonal isolate (data not shown).

† Anchorage-independent growth of HEK and BJ cell populations before (parental) and after (explanted) passage in a nude mouse.

10 population doublings, and no immortal clones arose spontaneously from these cultures after two months of culture, confirming earlier observations⁴. Oncogenic *ras* led to clear morphological transformation (data not shown) but had only a minor effect on the growth rate in monolayers of BJ fibroblasts expressing both large-T and *hTERT* (Fig. 3b).

We then ascertained the ability of these various cell types to grow in an anchorage-independent fashion, one of the hallmarks of the tumorigenic state. We observed efficient colony formation in soft agar only with cells expressing the combination of large-T, *ras* and *hTERT* (Table 1). Although occasional colonies were seen with cells expressing only large-T and *ras*, these were both significantly less numerous as well as markedly smaller in size. Furthermore, when cells expressing only large-T, large-T and *ras*, or large-T and *hTERT* were introduced into immunodeficient nude mice, no tumours were observed even after three months of observation, although the telomerase-expressing human breast cancer cell lines BT549 or SW613, used as controls, readily formed tumours in this assay (Table 2).

In marked contrast, when cells expressing large-T, *ras* and *hTERT* were introduced into nude mice, rapidly growing tumours were repeatedly observed with high efficiency (Table 2, Fig. 4c). Histologically, these converted HEK cells formed malignant tumour nodules composed of cells with cytoplasmic vacuoles, whereas the BJ fibroblasts were transformed into large, malignant, undifferentiated, spindle to epithelioid-shaped cells (data not shown). Thus, the ectopic expression of these three genetic elements seemed to be

sufficient to confer tumorigenic potential on both HEK cells and human fibroblasts.

But it remained possible that additional genetic alterations were required beyond these three changes for these cells to become tumorigenic. Several observations made such a scenario unlikely. We observed no lag in the outgrowth of tumours after injection of these cells into mice (Fig. 4c). Analysis of retroviral integration sites by Southern blotting revealed that the cell populations expressing large-T, *hTERT* and *ras* were polyclonal both before injection into nude mice and after recovery from them (Fig. 4b). Explanted tumour cells were morphologically indistinguishable (data not shown), had similar telomere lengths (Fig. 2b), grew *in vitro* at the same rate (Fig. 4a) and formed similar numbers of anchorage-independent colonies to the parental cells (Table 1, Experiment 2). Furthermore, when these explanted tumour cells were reinoculated into nude mice, they formed tumours with similar kinetics to those seen after injection of the parental cells (Fig. 4c). Taken together, these findings suggest that the tumorigenic growth exhibited by these cells is not the consequence of additional, rare stochastic events occurring *in vivo* after inoculation of these cells.

We conclude that the ectopic expression of a defined set of genes, specifically large-T, *ras* and *hTERT*, suffices to convert normal human cells into tumorigenic cells. These results begin to define the biochemical pathways that must be disrupted to create tumorigenic human cells from normal mesenchymal or epithelial precursors. This number is likely to involve changes in at least four distinct signalling pathways, minimally one more than is needed to transform rodent cells. In addition to changes required in the mitogen-response pathway activated by *ras* and in the telomere maintenance pathway represented by *hTERT*, large-T perturbs at least two distinct cellular control pathways through its ability to bind and

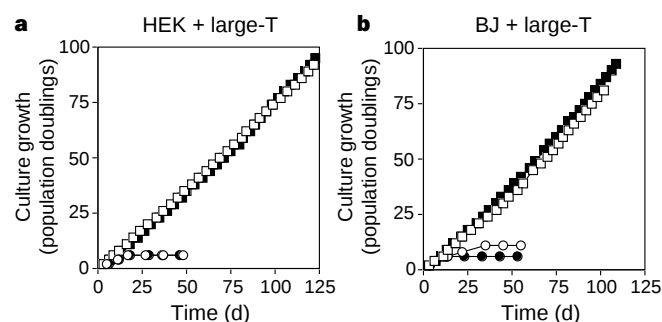


Figure 3 Expression of *hTERT* immortalizes large-T-expressing HEK and BJ cells. Growth of HEK (a) and BJ (b) cell populations is shown. Cells lacking *hTERT* (circles) entered crisis and could no longer be passaged. Filled symbols represent cells that also expressed oncogenic *ras*. Squares indicate cells expressing *hTERT*.

Table 2 Formation of tumours in immunodeficient nude mice

Cells	Genotype	Number of tumors/number of injections
HEK + LT	hTERT ⁻ , Ras ⁻	0/6
	hTERT ⁻ , Ras ⁺	0/6
	hTERT ⁺ , Ras ⁻	0/8
	hTERT ⁺ , Ras ⁺	8/8
BJ + LT	hTERT ⁻ , Ras ⁻	0/4
	hTERT ⁻ , Ras ⁺	0/4
	hTERT ⁺ , Ras ⁻	0/4
	hTERT ⁺ , Ras ⁺	4/4
BT549		8/8
SW613		8/8

Results are shown for polyclonal populations. Analysis of two individual clones derived from each population gave identical results; 2 × 10⁶ cells were injected in each experiment. LT, large-T.

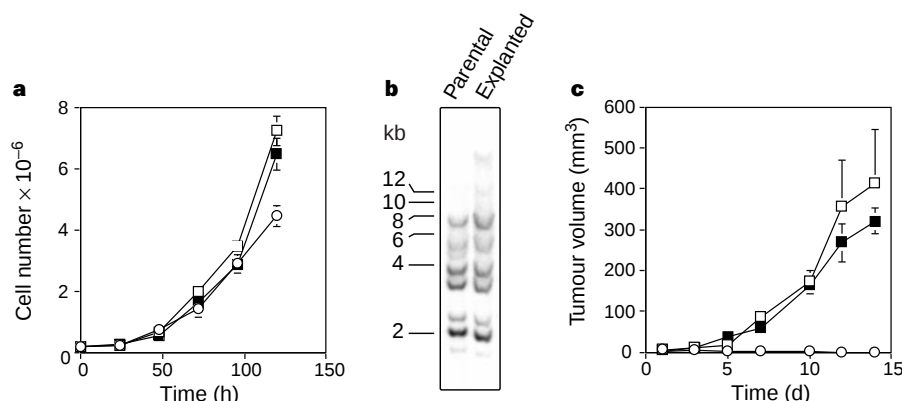


Figure 4 Growth properties and clonality of tumorigenic cells. **a**, Explanted or parental HEK cells expressing large-T, *ras* and *hTERT* grew at equivalent rates in culture. HEK cells expressing large-T and *hTERT* but lacking (open circles) or expressing (open squares) *ras* were compared with explanted tumour cells (filled squares). **b**, Comparison of retroviral integration sites in parental or explanted HEK cells expressing large-T, *ras* and *hTERT* demonstrates that such cells are

polyclonal. An *EcoRV*–*Sall* fragment of the pBABE-hygro-hTERT retrovirus²¹ was used as a probe. The positions of size standards (kb) are indicated at the left. **c**, Explanted tumour cells grew at the same rate as parental HEK cells expressing large-T, *ras* and *hTERT* on reinoculation into nude mice. Symbols are the same as in **a**. Results are means $\pm$ s.d. for six experiments.

functionally inactivate the cellular pRB and p53 tumour-suppressor proteins. Others have recently reported that the introduction of *hTERT*, *ras* and the human papillomavirus E6 and E7 proteins (which, like large-T, inactivate p53 and pRB) together did not succeed in allowing normal human fibroblasts to grow in an anchorage-independent fashion²². Because large-T is also known to perturb other cellular targets^{23,24}, the possibility remains that one or more of these other pathways must also be disrupted for tumorigenic conversion to occur. Nevertheless, our findings indicate that only in the context of *hTERT* expression do alterations in these other cellular pathways lead to tumorigenesis in human cells.

It is now highly likely that telomere maintenance, achieved either by *hTERT* expression or by an alternative mechanism²⁵, contributes directly to oncogenesis by allowing pre-cancerous cells to proliferate beyond the number of replicative doublings allotted to their normal precursors. Although the expression of telomerase does not by itself lead to a tumorigenic phenotype (Table 2)^{22,26}, telomere maintenance facilitated by *hTERT* expression *in vivo* might cooperate with additional oncogenic mutations to create a malignantly transformed clone. We suspect that the small, abortive soft agar colonies seen here in the absence of ectopic *hTERT* expression are one manifestation of replicative mortality that occurs in the absence of telomere maintenance. The present observations support the notion that telomere maintenance is essential for the formation of human tumour cells. We suggest that identical rules will be found to apply to autochthonously arising human tumour cells. □

Methods

Generation of cell lines. To create amphotropic retroviruses, Phoenix cells were transfected with pBABE-hygro-hTERT²¹ and pBABE-puro-*ras*-V12 (a gift from S. Lowe) by calcium-phosphate precipitation. These ecotropic retroviral supernatants and supernatants collected from the large-T retroviral producer line ψ_2 : simian virus 40 (ref. 27) were used to infect the amphotropic packaging cell line PT67 (Clontech Laboratories). The resulting supernatants were used to infect HEK (HA1)²¹ cells or foreskin fibroblasts (BJ). Retroviral constructs were introduced serially; drug selection was used to purify cell populations between infections. Cells were selected in hygromycin (100 μ g ml⁻¹, 7 d), puromycin (0.5 μ g ml⁻¹, 2–3 days) or neomycin (400 μ g ml⁻¹, 7 days), respectively. Retroviral vectors carrying only drug resistance genes were used as controls. In all cases, the point at which a culture reached confluence in a 10-cm culture dish after the last viral infection was designated population doubling 0; this point represents at least 70 population doublings from the original primary culture. Cells were considered to have entered crisis when they could no longer be passaged and exhibited widespread cell death.

Telomerase assays, polymerase chain reaction with reverse transcription

(RT-PCR), telomere analysis and immunoblotting. Cellular extracts were assayed for telomerase activity with a PCR-based telomeric repeat amplification protocol (TRAP) assay²⁸. For RT-PCR, total cellular RNA was prepared from cells by using RNazol (Tel Test B). In each reaction, 100 ng of total RNA was used with primers specific for retrovirally encoded *hTERT* (5′-GACACA-CATTCACAGGTCG-3′ and 5′-GACTGACACCGTGTCACCTAC-3′) or primers specific for human GADPH (5′-GAGAGACCCTCACTGCTG-3′ and 5′-GATGGTACATGACAAGGTGC-3′) with the RTth kit (Perkin Elmer). The RT reaction was performed for 10 min at 70 °C followed by 30 cycles of PCR (94 °C, 45 s; 60 °C, 45 s; 72 °C, 90 s). Telomere length was measured by hybridizing a ³²P-labelled telomeric (CCCTAA)₃ probe to *Hinf*I- and *Ras*I-digested genomic DNA. Immunoblotting of total cellular extracts (75 μ g) was performed with the rat anti-Ras antibody 259 and the mouse anti-LT antibody Pab 101 (Santa Cruz Biotechnology).

Analysis of retroviral integration sites. Genomic DNA (15 μ g) from parental and explanted tumour cells was digested with *Eco*RI and *Bam*HI and hybridized to a ³²P-labelled *Eco*RV–*Sall* fragment of the *hTERT* cDNA under conditions that did not permit the detection of genomic fragments of *hTERT*.

Soft agar assays. Soft agar assays were performed as described²⁹. Cultures were coded and scored in a blinded fashion by a second observer.

Tumorigenicity assays. Immunodeficient mice (Balb/c–ByJ–Hfh11^{nu}; Jackson Laboratory) were maintained in pathogen-free conditions. Mice were γ -irradiated with 400 rad before injection to suppress natural killer cells³⁰; tumours also formed in mice that had not been irradiated, but with a 1–2 week latency period. Cells (2×10^6) were injected subcutaneously into mice anaesthetized with Metofane; tumours were measured every 2–3 d. Mice were when tumours exceeded 1 cm. In experiments where cells were reisolated, tumours were removed surgically, minced, incubated in a dilute (0.15%) solution of collagenase for 2 h, washed, and placed in culture. Tumour volume was calculated with the formula $4/3\pi r^3$, where r is the radius of the tumour.

Received 2 February; accepted 18 May 1999.

- Land, H., Parada, L. F. & Weinberg, R. A. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature* **304**, 596–602 (1983).
- Ruley, H. Adenovirus early region 1A enables viral and cellular transforming genes to transform primary cells in culture. *Nature* **304**, 602–606 (1983).
- Sager, R. Senescence as a mode of tumor suppression. *Environ. Health Perspect.* **93**, 59–62 (1991).
- O'Brien, W., Stenman, G. & Sager, R. Suppression of tumor growth by senescence in virally transformed human fibroblasts. *Proc. Natl Acad. Sci. USA* **83**, 8659–8663 (1986).
- Stevenson, M. & Volsky, D. J. Activated *v-myc* and *v-ras* oncogenes do not transform normal human lymphocytes. *Mol. Cell. Biol.* **6**, 3410–3417 (1986).
- Kang, J.-S. *et al.* Involvement of tyrosine phosphorylation of p185-erbB in tumorigenicity induced by X-rays and the *neu* oncogene in human breast epithelial cells. *Mol. Carcinogen.* **21**, 225–233 (1998).
- Yakum, G. H. *et al.* Transformation of human bronchial epithelial cells transfected by the Harvey *ras* oncogene. *Science* **227**, 1174–1179 (1985).
- Rhim, J. S. *et al.* Neoplastic transformation of human epidermal keratinocytes by AD12-SV40 and Kirsten sarcoma viruses. *Science* **227**, 1250–1252 (1985).
- Hurlin, P. J., Maher, V. M. & McCormick, J. J. Malignant transformation of human fibroblasts caused by expression of a transfected T24 HRAS oncogene. *Proc. Natl Acad. Sci. USA* **86**, 187–191 (1989).

10. Burger, A. M. *et al.* Effect of oncogene expression on telomerase activation and telomere length in human endothelial, fibroblast and prostate epithelial cells. *Int. J. Oncol.* **13**, 1043–1048 (1998).
11. Flore, A. *et al.* Transformation of primary human endothelial cells by Kaposi's sarcoma-associated herpesvirus. *Nature* **394**, 588–592 (1998).
12. Nakamura, T. M. & Cech, T. R. Reversing time: origin of telomerase. *Cell* **92**, 587–590 (1998).
13. Kipling, D. Telomere structure and telomerase expression during mouse development and tumorigenesis. *Eur. J. Cancer* **33**, 792–800 (1997).
14. Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011–2015 (1994).
15. Shay, J. W. & Bacchetti, S. A survey of telomerase activity in human cancer. *Eur. J. Cancer* **33**, 787–791 (1997).
16. Harley, C. B. *et al.* Telomerase, cell immortality, and cancer. *Cold Spring Harb. Symp. Quant. Biol.* **59**, 307–315 (1994).
17. Bodnar, A. G. *et al.* Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349–352 (1998).
18. Kiyono, T. *et al.* Both Rb/p16INK4a inactivation and telomerase activity are required to immortalize human epithelial cells. *Nature* **396**, 84–88 (1998).
19. Lustig, A. J. Crisis intervention: the role of telomerase. *Proc. Natl Acad. Sci. USA* **96**, 3339–3341 (1999).
20. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic *ras* provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* **88**, 593–602 (1997).
21. Counter, C. M. *et al.* Dissociation among *in vitro* telomerase activity, telomere maintenance, and cellular immortalization. *Proc. Natl Acad. Sci. USA* **95**, 14723–14728 (1998).
22. Morales, C. P. *et al.* Absence of cancer-associated changes in human fibroblasts immortalized with telomerase. *Nature Genet.* **21**, 115–118 (1999).
23. Zalvide, J., Stubb, H. & DeCaprio, J. A. The J domain of simian virus 40 large T antigen is required to functionally inactivate RB family proteins. *Mol. Cell. Biol.* **18**, 1408–1415 (1998).
24. Damania, B., Mital, R. & Alwine, J. C. Simian virus 40 large T antigen interacts with human TFIIB-related factor and small nuclear RNA-activating protein complex for transcriptional activation of TATA-containing polymerase III promoter. *Mol. Cell. Biol.* **18**, 1331–1338 (1998).
25. Reddel, R. R., Bryan, T. M. & Murnane, J. P. Immortalized cells with no detectable telomerase activity. A review. *Biochemistry (Mosc.)* **62**, 1254–1262 (1997).
26. Jiang, X.-R. *et al.* Telomerase expression in human somatic cells does not induce changes associated with a transformed phenotype. *Nature Genet.* **21**, 111–114 (1999).
27. Jat, P. S., Cepko, C. L., Mulligan, R. C. & Sharp, P. A. Recombinant retroviruses encoding simian virus 40 large T antigen and polyomavirus large and middle T antigens. *Mol. Cell. Biol.* **6**, 1204–1217 (1986).
28. Kim, N. W. & Wu, F. Advances in quantification and characterization of telomerase activity by the telomeric repeat amplification protocol (TRAP). *Nucleic Acids Res.* **25**, 2595–2597 (1997).
29. Cifone, M. A. & Fidler, I. J. Correlation of patterns of anchorage-independent growth with *in vivo* behavior of cells from a murine fibrosarcoma. *Proc. Natl Acad. Sci. USA* **77**, 1039–1043 (1980).
30. Feuer, G. *et al.* Potential role of natural killer cells in controlling tumorigenesis by human T-cell leukemia viruses. *J. Virol.* **69**, 1328–1333 (1995).

Acknowledgements. We thank M. Fleming for interpretation of tumour histology, J. Smith for the gift of early passage BJ fibroblasts, and S. Dessain, B. Elenbaas, D. Fruman, P. Steiner, S. Stewart and the members of Weinberg laboratory for helpful discussions and review of the manuscript. This work was supported in part by Merck and Co. (R.A.W.), the US NCI (R.A.W., A.S.L.), a Damon Runyon–Walter Winchell Cancer Research Foundation Postdoctoral Fellowship (W.C.H.), and a Human Frontiers Postdoctoral Fellowship (R.L.B.). C.M.C. is a Whitehead Scholar; W.C.H. is a Herman and Margaret Sokol postdoctoral fellow. R.A.W. is an American Cancer Society Research Professor and a Daniel K. Ludwig Cancer Research Professor.

Correspondence and requests for material should be addressed to R.A.W. (e-mail: weinberg@wi.mit.edu).

Essential role for oncogenic Ras in tumour maintenance

Lynda Chin^{†‡}, Alice Tam^{†‡§}, Jason Pomerantz[§], Michelle Wong[§], Jocelyn Holash^{||}, Nabeel Bardeesy[†], Qiong Shen[†], Ronan O'Hagan[†], Joe Pantginis^{||}, Hao Zhou^{||}, James W. Horner II[†], Carlos Cordon-Cardo[†], George D. Yancopoulos^{||} & Ronald A. DePinho^{†#}

[†]Adult Oncology, Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA

[‡]Department of Dermatology, Harvard Medical School, Boston, Massachusetts 02115, USA

[§]Department of Medicine and Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

^{||}Albert Einstein College of Medicine, Bronx, New York 10461, USA

[¶]Memorial Sloan Kettering Cancer Center, New York, New York 10021, USA

^{||}Regeneron Pharmaceuticals, Tarrytown, New York 10591, USA

* These authors contributed equally to this work

Advanced malignancy in tumours represents the phenotypic endpoint of successive genetic lesions that affect the function and regulation of oncogenes and tumour-suppressor genes¹. The established tumour is maintained through complex and poorly understood host–tumour interactions that guide processes such as angiogenesis and immune sequestration. The many different genetic alterations that accompany tumour genesis raise ques-

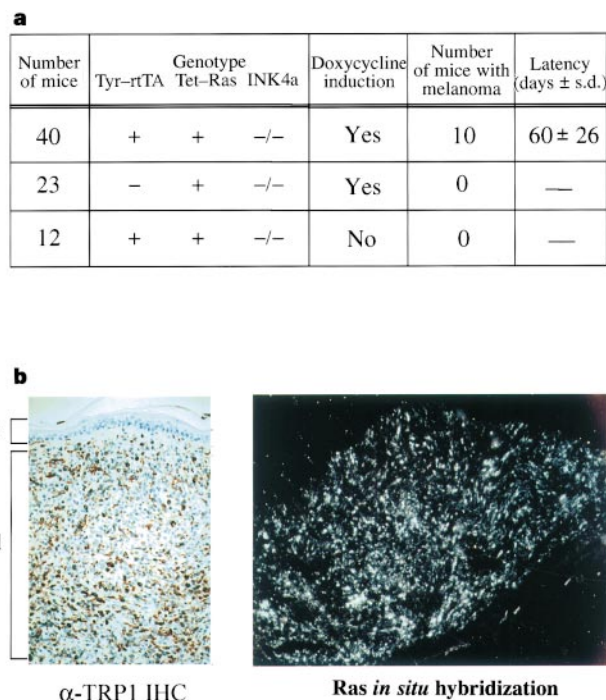


Figure 1 Inducible Tyr/Tet-Ras transgenic mice on INK4a-deficient background developed cutaneous melanomas. **a**, Summary of tumour incidence in the Tyr/Tet-Ras-INK4a-null colony and impact of doxycycline treatment. **b**, Left: anti-TRP1 staining of a primary cutaneous melanoma. Note strong immunoreactivity in dermal nodule with overlying intact epidermis; **e**, epidermis; **d**, dermis. Right: *in situ* hybridization for H-Ras^{V12G} transcript in a primary melanoma nodule.

tions as to whether experimental cancer-promoting mutations remain relevant during tumour maintenance. Here we show that melanoma genesis and maintenance are strictly dependent upon expression of H-Ras^{V12G} in a doxycycline-inducible H-Ras^{V12G} mouse melanoma model null for the tumour suppressor INK4a. Withdrawal of doxycycline and H-Ras^{V12G} down-regulation resulted in clinical and histological regression of primary and explanted tumours. The initial stages of regression involved marked apoptosis in the tumour cells and host-derived endothelial cells. Although the regulation of vascular endothelial growth factor (VEGF) was found to be Ras-dependent *in vitro*, the failure of persistent endogenous and enforced VEGF expression to sustain tumour viability indicates that the tumour-maintaining actions of activated Ras extend beyond the regulation of VEGF expression *in vivo*. Our results provide genetic evidence that H-Ras^{V12G} is important in both the genesis and maintenance of solid tumours.

To develop a cancer model in which dominantly acting oncoproteins are somatically regulated *in vivo*, transgenic mouse lines harbouring the reverse tetracycline transactivator² under the control of the tyrosinase gene promoter–enhancer elements (designated Tyr-rtTA) and another containing the H-Ras^{V12G} open reading frame driven by a minimal promoter containing multimerized tet-operons^{2,3} (designated Tet-Ras) were inter-crossed with INK4a^{-/-} mice to generate cohorts of single and double transgenic mice (designated Tyr/Tet-Ras) that were homozygous null for INK4a. Upon weaning, a subset of single and double transgenic INK4a^{-/-} mice were given doxycycline in their drinking water⁴. In the doxycycline-treated group, 10 out of 40 Tyr/Tet-Ras INK4a^{-/-} mice developed melanomas with an average latency of 60 days (Fig. 1a). In contrast, the untreated Tyr/Tet-Ras INK4a^{-/-} mice (*n* = 12) or treated Tet-Ras INK4a^{-/-} (*n* = 23) did not develop melanomas. The Tyr/Tet-Ras INK4a^{-/-} melanomas shared all of the macroscopic features of the constitutive Tyr-Ras INK4a^{-/-}

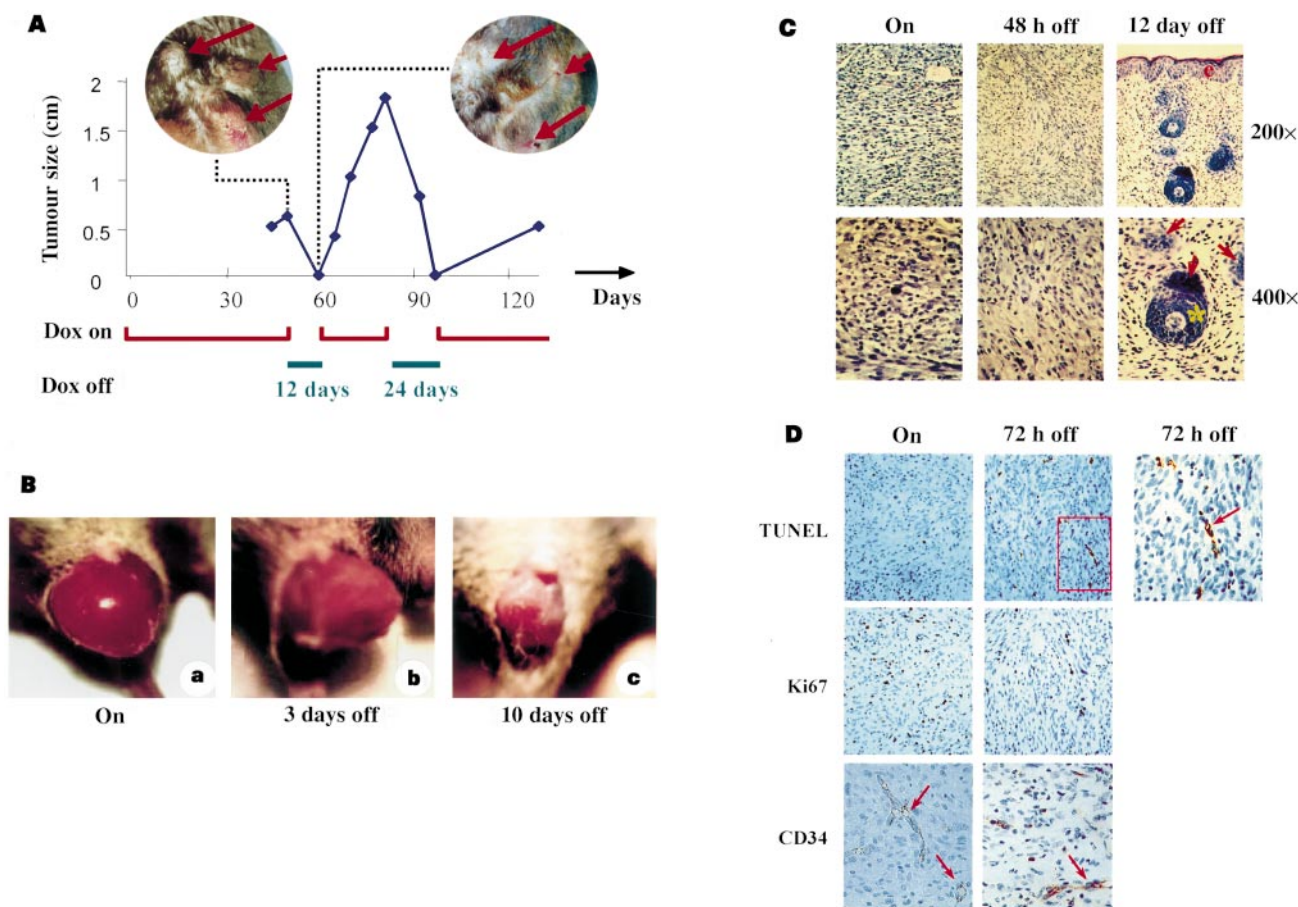


Figure 2 Activated Ras expression is necessary to maintain growth of established cutaneous melanomas *in vivo*. **A**, Tumour from R348 (Tet-Ras founder line 65) was measured at intervals and the longest dimension was plotted against time. Time 0 represents initial induction with doxycycline. Red brackets, periods during which the animal was on doxycycline induction; blue lines, periods without doxycycline administration. Insets: Representative photographs of R348 tumour taken at times shown (on and 12 days off) during the first cycle of doxycycline treatment. All primary melanomas ($n = 10$) regressed rapidly following doxycycline withdrawal, although two of the largest tumours recurred at the site as slow-growing tumours of indeterminate type (see text). Two fully regressed tumour-bearing mice were re-induced to form large tumours, one of which regressed fully after doxycycline withdrawal (shown here), and the other showed stable growth for 10 more days before being killed. **B**, Regression of

melanomas⁵, manifesting as amelanotic, invasive and highly vascular tumours—features that are reminiscent of nodular-type melanoma in humans (for example, see Fig. 2A, B). Histological examination revealed a spindle morphology with anaplastic and pleiomorphic cytology (Fig. 2C), strong immunoreactivity to the early melanocyte-specific marker tyrosinase-related-protein-1 (TRP-1)⁶, and robust H-Ras^{V12G} expression and activity in tumours and tumour-derived cell lines in culture (Figs 1b and 4a).

Although these and previous results^{5,7} confirm a causal role for H-Ras^{V12G} in melanoma development, the ability to regulate H-Ras^{V12G} activity in these melanomas allowed us to determine the role of activated Ras in tumour maintenance. To this end, different doxycycline-treated Tyr/Tet-Ras INK4a^{-/-} mice with one or more independent primary melanomas, ranging in size from 0.5 to 1.5 cm in diameter, had doxycycline administration withdrawn (Fig. 2A, B). Following an initial loss of tumour erythema (Fig. 2B, compare b to a), these large tumours regressed to barely detectable or undetectable lesions with only residual scattered tumour foci on microscopic examination by day 14 of doxycycline withdrawal (Fig. 2C, compare right panel with left). Consistent with the

presence of residual microscopic disease, re-administration of doxycycline resulted in the rapid recurrence of tumours at previous tumour sites (Fig. 2A). This model may be useful in the development of anti-oncologics directed towards minimal residual disease. All primary melanomas ($n = 10$) exhibited regression upon doxycycline withdrawal; however, doxycycline-independent tumours can persist or re-emerge ($n = 3$), particularly with large tumour burdens (Fig. 2). These later tumours were whitish, exhibited a less anaplastic spindle cytology, failed to express the H-Ras^{V12G} transgene (by western blot analysis) and were TRP-1 negative (data not shown). These data indicate that these doxycycline-independent tumours are either a different cancer type, such as fibrosarcoma, or Ras-independent melanomas.

Although these results and those of others⁸ confirm a role for Ras in tumorigenesis, this model system allows us to investigate mechanisms underlying tumour regression *in vivo*. We first assessed the impact of doxycycline withdrawal in different Tyr/Tet-Ras INK4a^{-/-} tumour-derived cell lines in cell culture. Melanoma cell lines were established from primary tumours in doxycycline-supplemented medium and were shown to express H-Ras^{V12G} messenger

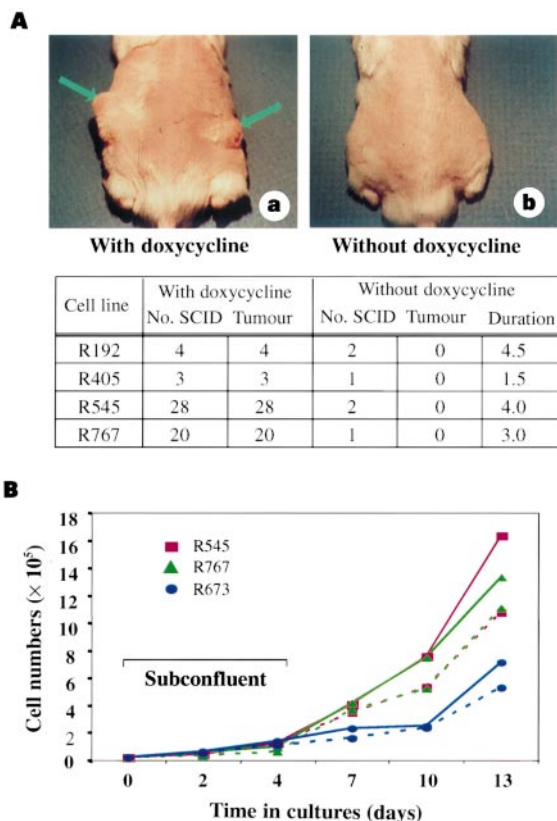


Figure 3 *In vitro* and *in vivo* behaviour of melanoma cells in the presence and absence of doxycycline. **A**, Photomicrograph of SCID mice maintained on (a) and off (b) doxycycline after subcutaneous injection of R192 cells. Arrows, subcutaneous tumours. Similar experiments performed with multiple independently derived melanoma cell lines are summarized in the table. 'Duration' refers to the period of observation in months. **B**, Growth-curves of three independently derived melanoma cell lines with and without doxycycline supplementation in medium (see Methods). Solid lines, doxycycline-treated; dotted lines, doxycycline-free. Cultures were subconfluent during the first 5 days.

ger RNA, protein, and activity in a doxycycline-dependent manner (Fig. 4a and data not shown). The derivative melanoma cell lines remained doxycycline-responsive, as shown by their strict doxycycline-dependency for tumorigenic potential in severe-combined immunodeficient (SCID) mice. Specifically, subcutaneous injection of purified Tyr/Tel-Ras INK4a^{-/-} melanoma cell lines yielded tumours in SCID mice treated with doxycycline with 100% efficiency, whereas those mice not receiving doxycycline remained tumour-free after many months of observation (Fig. 3A). In cell culture, the presence or absence of doxycycline did not significantly influence the subconfluent growth rates of multiple independently derived lines (Fig. 3B). Similarly, the presence of doxycycline did not confer enhanced growth potential in low-serum conditions (0.5 and 1% FCS; data not shown). However, the presence of doxycycline slightly enhanced growth in dense cultures, indicating that *in vivo* Ras expression may provide a continued cell-autonomous advantage for tumour growth. On the other hand, the data is also consistent with the possibility that tumour regression following doxycycline withdrawal may reflect a continued requirement for Ras in promoting and sustaining non-cell-autonomous tumour-host interactions that are essential for tumour growth and maintenance.

One hypothesis supporting the latter possibility is that continued expression of Ras confers on the tumour cells the capacity to evade the host immune response⁹; thus, tumour regression is due to onset of immune rejection by the host. To examine the contribution of an intact immune system to tumour regression, we determined the regression kinetics of tumours derived from purified Tyr/Tet-Ras

INK4a^{-/-} melanoma cell lines in SCID tumour explants. Upon doxycycline withdrawal, established tumours showed rapid reductions in tumour size accompanied by histological, proliferative and apoptotic profiles that were indistinguishable from those of the primary tumours in immune-competent hosts (data not shown). As SCID mice lack functioning B and T lymphocytes (although they have functional natural killer cells and macrophages), these data are consistent with the view that the immune system is not critical during the initial phase in which the bulk of the tumour burden is eliminated.

Another important aspect of tumour reliance on host interactions involves angiogenic support. In fact, examination of regressing tumours indicated that vascular regression might even precede frank tumour loss. We performed serial biopsies of the same primary Tyr/Tet-Ras INK4a^{-/-} tumours following doxycycline withdrawal at 0, 24, 48 and 72 h. Tumour-bearing mice from Tet-Ras founder lines 65 and 72 were examined (for example, R348 and R405 from line 65 and 72, respectively) and yielded identical phenotypes. Upon microscopic examination, a morphological transition from severe anaplasia to a more organized spindle cytoarchitecture occurred by 48 h after doxycycline withdrawal (R348 tumour shown in Fig. 2C). A slight decline in proliferative index and a marked increase in apoptosis (R405 tumour shown in Fig. 2D) accompanied these histological changes. In many primary tumours that we examined, the apoptotic response was evident as early as 24 h after doxycycline withdrawal (data not shown). Most notably, many TUNEL-positive cells were found in close proximity to small tumour vessels, specifically the cells comprising the vessel wall (Fig. 2D). Vessel-associated apoptosis coincided with a marked decrease in immunoreactivity for CD34 and CD31, two classical endothelial-cell markers^{10,11} (Fig. 2D; CD31 not shown).

The significant apoptotic rate of cells lining tumour vessels suggested that continued activated-Ras expression may be required for the critical host-tumour symbiotic interaction that sustains stable tumour vasculature. Consistent with this, previous cell-culture data have shown that oncogenic K- and H-Ras^{V12G} can stimulate VEGF expression¹²⁻¹⁴. To determine the role, if any, of VEGF in Ras-dependent tumour maintenance, Ras activity (as measured by the Raf-GST pull-down assay) and VEGF protein level (see Methods) were serially determined in cell culture and tumour explants following doxycycline withdrawal. In cell culture, a decline in H-Ras^{V12G} activity correlated with downregulation of VEGF expression (Fig. 4a, b). In the tumour explants, although a decline in Ras activity led to an eventual decrease in VEGF during later stages of tumour regression (after day 3; Fig. 4a, b), early regression and endothelial-cell apoptosis occurred despite sustained VEGF expression (Fig. 4c, d), presumably due to hypoxic/anoxic stimuli¹⁵⁻¹⁸. The contrasting VEGF kinetics indicated that tumour regression *in vivo* is not mediated by the loss of Ras-stimulated VEGF expression and that high VEGF levels are insufficient to maintain tumour viability in the absence of H-Ras^{V12G} expression. To examine this issue more directly, the doxycycline-responsive R545 melanoma cell line was engineered to constitutively express either GFP (green fluorescent protein) or VEGF with GFP. The VEGF-transduced cell lines expressed on average 10 times more VEGF than the GFP-transduced control in the absence of doxycycline (10 ng ml⁻¹ versus <1 ng ml⁻¹ by ELISA). Although the VEGF-transduced cells showed faster tumour growth, withdrawal of doxycycline resulted in regression of these tumours, despite enforced VEGF expression (Fig. 4c). Furthermore, as in primary tumours and SCID tumours derived from the parental untransduced melanoma lines, this tumour regression was associated with marked activation of apoptosis in tumour cells and in host endothelial cells (Fig. 4d).

Together, these data show that, although activated Ras regulates VEGF expression in tumour cells as previously reported^{13,19-21}, complex and poorly understood Ras-independent mechanisms

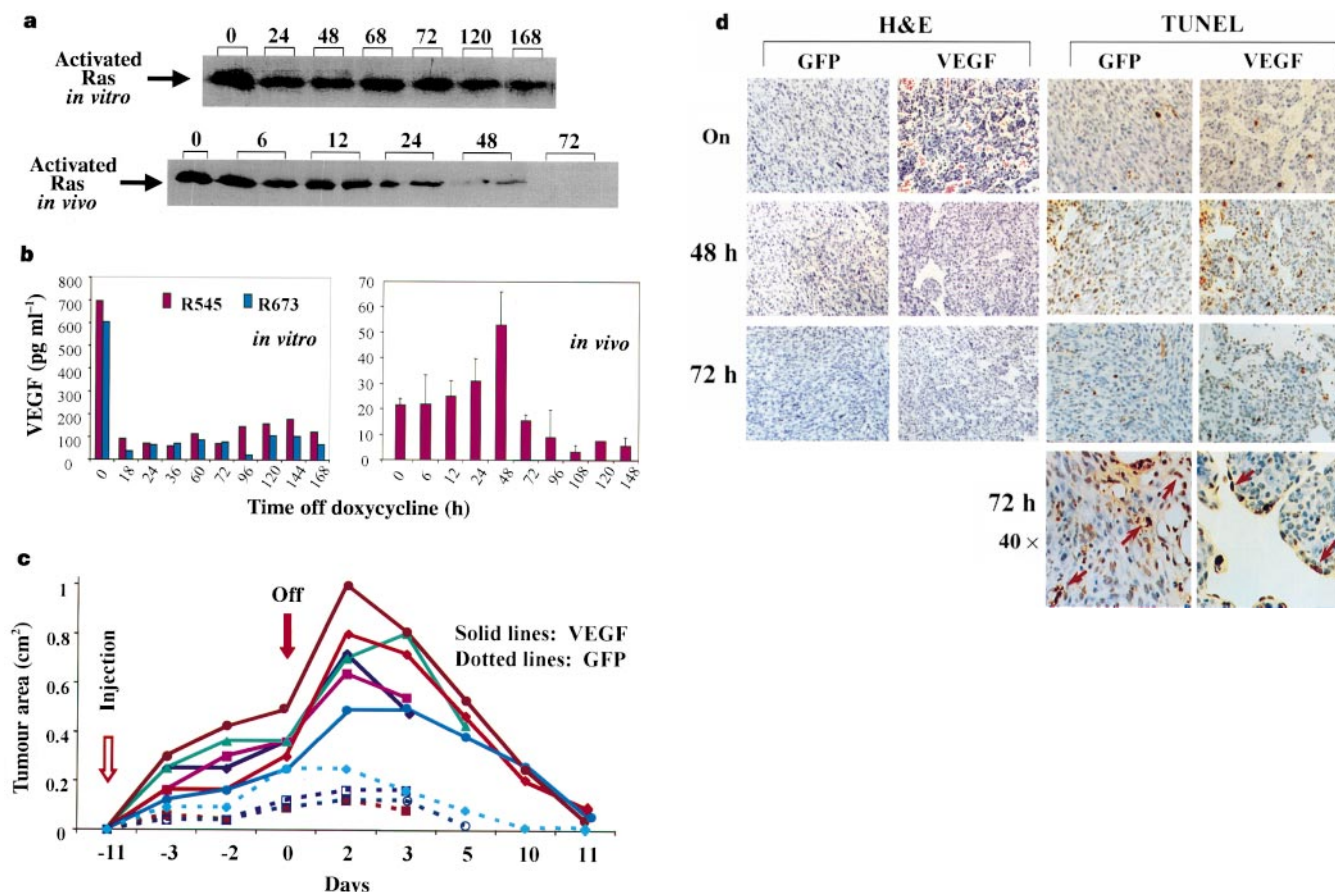


Figure 4 VEGF is not sufficient to sustain tumour viability following doxycycline withdrawal. **a**, Ras activity after withdrawal from doxycycline *in vivo* and *in vitro*. R545 melanoma cells in cultures or R545-derived SCID tumours were subjected to Raf-GST pull-down assay (see Methods) to determine Ras activity at indicated time after doxycycline withdrawal. For *in vivo* assay, two independently derived samples were assayed except for the 0 h time-point. **b**, Endogenous VEGF protein in cultured R545 and R673 cells (left) and SCID-derived R545 tumours (right). **c**, Enforced VEGF expression in R545 melanoma cells is not sufficient to block

tumour regression in SCID mice after doxycycline withdrawal. Transduced cells were injected on day -11 and doxycycline was withdrawn from drinking water on day 0. **d**, Haematoxylin and eosin (H&E) and TUNEL staining of SCID tumours from **c**. Note large ectatic vascular spaces in VEGF-transduced tumours. Increase in number of TUNEL-positive nuclei (both tumour and endothelial cells) was evident both 48 h and 72 h after doxycycline withdrawal in tumours with or without enforced VEGF expression. Arrows, TUNEL-positive endothelial cells lining vascular spaces.

contribute significantly to the regulation of VEGF in established tumours upon doxycycline withdrawal. Correspondingly, the fact that initiation of tumour regression and associated vascular collapse take place despite either robust endogenous or retrovirally enforced VEGF expression indicates that VEGF is not sufficient for tumour maintenance, consistent with previous findings¹⁴, and that the role of activated Ras in tumour maintenance extends beyond the regulation of VEGF expression. Thus, the *in vivo* model reported here will serve to uncover additional aspects of the complex homotypic and heterotypic interactions between host and tumour cells that are essential for the maintenance of fully formed solid tumours. Whereas an improved understanding of such interactions could stimulate the development of therapeutics directed towards the less plastic host-cell compartment, it is notable that despite the occurrence of many genetic changes in these tumours, effective anti-Ras therapeutics would be predicted to hold significant promise in anticancer therapy. □

Methods

Production of the transgenic mice. The reverse tetracycline transactivator (rtTA) is a 1,050-base pair (bp) *EcoRI/BamHI* fragment isolated from pUHD172-1neo². The Tet promoter contains an *XhoI/EcoRI* fragment of the cytomegalovirus minimal promoter linked to the tet operator sequences². The tyrosinase enhancer/promoter and the H-Ras^{Val12} transgene were as described^{13,5,22}. We generated multiple founder lines for both transgenes at the expected frequencies. We used one activator line (Tyr/rtTA, line 37) and two independent reporter lines (Tet-RAS, lines 65 and 72) for these studies.

Primary tumours, derivative cell lines and SCID explant tumours. Transgenic mice were fed doxycycline drinking water (2 mg ml⁻¹ in sucrose water) and observed for spontaneous tumour development. Primary tumours were adapted to culture by mechanical mincing with sterilized razor blades and brief trypsinization, and maintained in RPMI medium containing 10% serum and supplemented with doxycycline (2 µg ml⁻¹ medium)⁴. For the SCID explant tumours, 2–5 × 10⁶ established melanoma cells were injected subcutaneously into the flanks of adult SCID mice maintained on doxycycline or regular drinking water. These cell lines were passaged sufficiently to insure elimination of immunocytes from the original host.

RNA and protein analysis. Total RNAs were isolated from tumour tissues and cells using Trizol reagents (Gibco BRL). For H-Ras^{Val12}, a 700-bp *SalI* complementary DNA fragment isolated from pBABE-Ras (a gift from S. Lowe) was used as a probe. For VEGF, a *SacI/KpnI* 600-bp ORF fragment was isolated from mouse VEGF plasmid (a gift from P. Maisonpierre). RNA *in situ* hybridization was performed as described²³ on tumour samples snap-frozen in OCT (Lab-Tek). Western blots of tumour or cell lysates (100 µl input) were run on 15% SDS-PAGE and probed with anti-Ras monoclonal antibodies (Oncogene). Ras activity was determined by a Ras activation assay (Upstate Biotechnology), following the manufacturer's protocol. VEGF protein levels were determined in duplicate by ELISA (R&D System) using 100 µg of protein (from tissue lysates or conditioned media). Microplates were read using the OPTiMax tunable microplate reader (Molecular Devices) and analysed with SoftMax Pro.

Histological analysis and immunohistochemistry. Tissue samples were formalin-fixed and paraffin-embedded. Anti-TRP1 staining using a TA99

monoclonal antibody⁶, TUNEL staining, Ki67 and anti-CD34 staining were performed as described^{5,11,24,25}.

Growth curves. Cells from each cell line were seeded at a density of 20,000 cells per well in a 12-well plate in medium with or without doxycycline. Medium was changed every 3 days for all samples. Duplicate wells were trypsinized, and cell numbers were counted by haemocytometer at indicated time points and plotted against time. Studies were conducted in media containing 10, 1 and 0.5% serum. Growth-curve determinations were performed in cells maintained on doxycycline before experiments as well as onset of cells that had been removed from doxycycline for 3 days before the experiments.

Enforced VEGF-expressing melanoma cell line. The R545 melanoma cell line was transduced with LZRSpBMN-IRES-GFP and LZRSpBMN-VEGF-IRES-GFP, and purified populations were established by FACS sorting²⁶. SCID explant experiments used transduced cell lines generated as described above. Tumour sizes were measured with tumour callipers in two dimensions and area plotted against time (in days).

Received 19 March; accepted 22 June 1999.

- Bishop, J. M. Molecular themes in oncogenesis. *Cell* **64**, 235–248 (1991).
- Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766–1769 (1995).
- Fasano, O., Taparowsky, E., Fiddes, J., Wigler, M. & Goldfarb, M. Sequence and structure of the coding region of the human H-ras-1 gene from T24 bladder carcinoma cells. *J. Mol. Appl. Genet.* **2**, 173–180 (1983).
- Kistner, A. *et al.* Doxycycline-mediated quantitative and tissue-specific control of gene expression in transgenic mice. *Proc. Natl Acad. Sci. USA* **93**, 10933–10938 (1996).
- Chin, L. *et al.* Cooperative effects of *INK4a* and *ras* in melanoma susceptibility *in vivo*. *Genes Dev.* **11**, 2822–2834 (1997).
- Thomson, T. M. *et al.* Differentiation antigens of melanocytes and melanoma: analysis of melanosome and cell surface markers of human pigmented cells with monoclonal antibodies. *J. Invest. Dermatol.* **90**, 459–466 (1988).
- Gause, P. R. *et al.* Chromosomal and genetic alterations of 7,12-dimethylbenz[a]anthracene-induced melanoma from TP-ras transgenic mice. *Mol. Carcin.* **20**, 78–87 (1997).
- Shirasawa, S., Furuse, M., Yokoyama, N. & Sasazuki, T. Altered growth of human colon cancer cell lines disrupted at activated Ki-ras. *Science* **260**, 85–88 (1993).
- Doherty, P. C., Tripp, R. A. & Sixbey, J. W. Evasion of host immune responses by tumours and viruses. *Ciba Found. Symp.* **187**, 245–256 (1994).
- Bird, I. N., Spragg, J. H., Ager, A. & Matthews, N. Studies of lymphocyte transendothelial migration: analysis of migrated cell phenotypes with regard to CD31 (PECAM-1), CD45RA and CD45RO. *Immunology* **80**, 553–560 (1993).
- Traweek, S. T., Kandalaf, P. L., Mehta, P. & Battifora, H. The human hematopoietic progenitor cell antigen (CD34) in vascular neoplasia. *Am. J. Clin. Pathol.* **96**, 25–31 (1991).
- Rak, J. *et al.* Oncogenes as inducers of tumor angiogenesis. *Cancer Metastasis Rev.* **14**, 263–277 (1995).
- Arbiser, J. L. *et al.* Oncogenic H-ras stimulates tumor angiogenesis by two distinct pathways. *Proc. Natl Acad. Sci. USA* **94**, 861–866 (1997).
- Okada, E. *et al.* Impact of oncogenes in tumor angiogenesis: mutant K-ras up-regulation of vascular endothelial growth factor/vascular permeability factor is necessary, but not sufficient for tumorigenicity of human colorectal carcinoma cells. *Proc. Natl Acad. Sci. USA* **95**, 3609–3614 (1998).
- Shweiki, D., Itin, A., Soffer, D. & Keshet, E. Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis. *Nature* **359**, 843–845 (1992).
- Mazure, N. M., Chen, E. Y., Yeh, P., Laderoute, K. R. & Giaccia, A. J. Oncogenic transformation and hypoxia synergistically act to modulate vascular endothelial growth factor expression. *Cancer Res.* **56**, 3436–3440 (1996).
- Goldberg, M. A. & Schneider, T. J. Similarities between the oxygen-sensing mechanisms regulating the expression of vascular endothelial growth factor and erythropoietin. *J. Biol. Chem.* **269**, 4355–4359 (1994).
- Mukhopadhyay, D. *et al.* Hypoxic induction of human vascular endothelial growth factor expression through c-Src activation. *Nature* **375**, 577–581 (1995).
- Rak, J. *et al.* Mutant ras oncogenes upregulate VEGF/VPF expression: implications for induction and inhibition of tumor angiogenesis. *Cancer Res.* **55**, 4575–4580 (1995).
- Grugel, S., Finkenzeller, G., Weindel, K., Barleon, B. & Marme, D. Both v-Ha-Ras and v-Raf stimulate expression of the vascular endothelial growth factor in NIH 3T3 cells. *J. Biol. Chem.* **270**, 25915–25919 (1995).
- Larcher, F. *et al.* Up-regulation of vascular endothelial growth factor/vascular permeability factor in mouse skin carcinogenesis correlates with malignant progression state and activated H-ras expression levels. *Cancer Res.* **56**, 5391–5396 (1996).
- Ganss, R., Montoliu, L., Monaghan, A. P. & Schutz, G. A cell-specific enhancer far upstream of the mouse tyrosinase gene confers high level and copy number-related expression in transgenic mice. *EMBO J.* **13**, 3083–3093 (1994).
- Valenzuela, D. M. *et al.* Receptor tyrosine kinase specific for the skeletal muscle lineage: expression in embryonic muscle, at the neuromuscular junction, and after injury. *Neuron* **15**, 573–584 (1995).
- Schreiber-Agus, N. *et al.* Role of Mx1 in ageing organ systems and the regulation of normal and neoplastic growth. *Nature* **393**, 483–487 (1998).
- Gavrieli, Y., Sherman, Y. & Ben-Sasson, S. A. Identification of programmed cell death *in situ* via specific labeling of nuclear DNA fragmentation. *J. Cell Biol.* **119**, 493–501 (1992).
- Cheng, L., Fu, J., Tsukamoto, A. & Hawley, R. G. Use of green fluorescent protein variants to monitor gene transfer and expression in mammalian cells. *Nature Biotech.* **14**, 606–609 (1996).

Acknowledgements. We thank G. Schutz for the tyrosinase promoter–enhancer elements; S. Jiao for tissue sample processing and immunohistochemistry; D. Compton for RNA *in situ* hybridization, and M. Russell for technical assistance. This work was supported by an NIH Mentored Clinician Scientist Award and the Harvard Skin Disease Center Grant (L.C.); by grants from the NIH (C.C.C. and R.A.D.) and from the DFCI Cancer Core (R.A.D. and L.C.). R.A.D. is an American Cancer Society Research Professor; A.T. and J.P. are HHMI Medical Student Fellows.

Correspondence and requests for materials should be addressed to R.A.D. (e-mail: ron_depinho@dfci.harvard.edu) or L.C. (e-mail: lynda_chin@dfci.harvard.edu).

Structure of the C-terminal domain of FliG, a component of the rotor in the bacterial flagellar motor

Scott A. Lloyd^{*‡}, Frank G. Whitby^{†‡}, David F. Blair^{*} & Christopher P. Hill[†]

^{*} Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, Utah 84112-0840, USA

[†] Department of Biochemistry, University of Utah, 50 North Medical Drive, Salt Lake City, Utah 84132, USA

[‡] These authors contributed equally to this work

Many motile species of bacteria are propelled by flagella, which are rigid helical filaments turned by rotary motors in the cell membrane^{1–3}. The motors are powered by the transmembrane gradient of protons or sodium ions. Although bacterial flagella contain many proteins, only three—MotA, MotB and FliG—participate closely in torque generation. MotA and MotB are ion-conducting membrane proteins that form the stator of the motor. FliG is a component of the rotor, present in about 25 copies per flagellum. It is composed of an amino-terminal domain that functions in flagellar assembly and a carboxy-terminal domain (FliG-C) that functions specifically in motor rotation. Here we report the crystal structure of FliG-C from the hyperthermophilic eubacterium *Thermotoga maritima*. Charged residues that are important for function, and which interact with the stator protein MotA^{4,5}, cluster along a prominent ridge on FliG-C. On the basis of the disposition of these residues, we present a hypothesis for the orientation of FliG-C domains in the flagellar motor, and propose a structural model for the part of the rotor that interacts with the stator.

We used FliG and FliG-C proteins from *Escherichia coli* for initial crystallization trials, because extensive mutational studies have been done using this species and the closely related species *Salmonella typhimurium*^{4,6}. Purified FliG and FliG-C from *E. coli* were polydisperse and failed to crystallize, however, which prompted us to study the protein from the hyperthermophile *T. maritima*⁷. The amino-acid sequence of *T. maritima* FliG-C (residues 209–335) is 46% identical to the protein from *E. coli* (Fig. 1), and charge is conserved at the functionally important positions. Null *fliG* mutants of *E. coli* are non-flagellate because FliG is required for flagellar assembly as well as rotation⁸. Both flagellar assembly and motility were restored to this mutant by a plasmid expressing a chimaeric protein consisting of residues 1–240 of *E. coli* FliG (the domain required for flagellar assembly) fused to residues 243–335 of *T. maritima* FliG (the domain required for rotation). Thus, the bulk of the FliG domain studied here is functionally interchangeable with the corresponding part of the *E. coli* protein. A plasmid that expressed full-length *T. maritima* FliG did not restore flagella to the mutant, presumably because the N-terminal domain of the *T. maritima* protein did not interact properly with other flagellar proteins of *E. coli*.

We overexpressed and purified full-length *T. maritima* FliG, but dynamic light scattering showed this to be highly polydisperse. A C-terminal domain consisting of residues 209–335 was found to be monodisperse. This domain was crystallized and its structure was determined by multiwavelength anomalous diffraction (MAD) and refined at 2.2 Å resolution to an *R*-factor (*R*_{free}) of 23.6% (30.4%) (Tables 1 and 2). The ordered residues (235–335) form a compact domain comprising six helices⁹ folded into a roughly triangular prism of approximate dimensions 38 Å × 30 Å × 27 Å (Figs 1 and

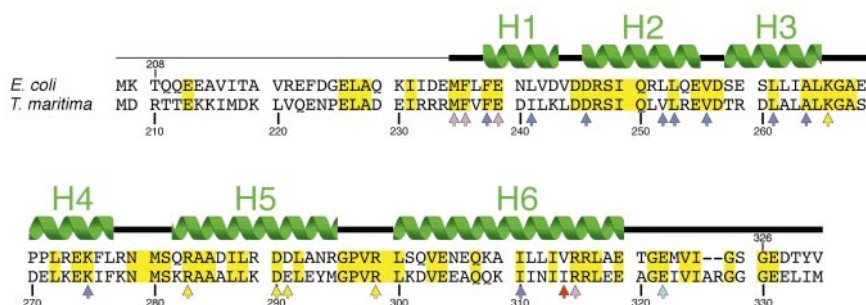


Figure 1 Alignment of *E. coli* and *T. maritima* FliG-C protein sequences. The initiating Met in the *T. maritima* sequence replaces a Leu in the native protein. Secondary-structure elements seen in the *T. maritima* protein are shown above. A thin line indicates disordered residues that were omitted from the model. Functionally important charged residues are indicated by yellow arrows, and positions of mutations discussed in the text are indicated by arrows coloured

according to phenotype: no motor rotation, blue; CW motor bias, mauve; weakened binding to FliM, red; suppressor of a mutation in MotB that alters the relationship of rotor and stator, turquoise. The mutations were isolated in *E. coli*^{4,14,16} and *S. typhimurium*⁸. The *T. maritima* fliG sequence was obtained from The Institute for Genomic Research at www.tigr.org. The *E. coli* sequence is from GenBank (accession number U46011).

2). The fold appears to be new as a search with the DALI¹⁰ server did not find any structurally similar proteins.

Several spontaneous mutations in FliG-C were found to prevent motor rotation while allowing assembly of normal-looking flagella, implicating this domain specifically in motor rotation⁶. Eight positions identified in those mutational studies contain hydrophobic residues (Phe 238, Ile 241, Val 252, Leu 253, Val 256, Leu 261, Ala 264 and Ile 310 in *T. maritima*) whose side chains together form much of the core of the domain (Fig. 2). (*T. maritima* residue numbers are used throughout this Letter.) The other two mutations that prevented motor rotation alter charged residues on the surface (Asp 246 and Lys 275) that form salt bridges or hydrogen bonds to nearby groups. Thus, as suggested previously⁶, the spontaneous mutations in FliG-C probably disrupt function by altering protein conformation rather than by changing residues of an active site.

Site-directed mutagenesis of FliG-C identified three charged residues of major importance for motor rotation (Arg 283, Asp 290 and Glu 291 in the *T. maritima* protein), and two others (Lys 266 and Arg 299) of secondary importance⁴. Mutant phenotypes indicate that these residues function redundantly, that charge

is their most important property, and that they interact with charged residues of the stator protein MotA^{4,5,11}. The functionally important charged residues are coloured according to atom type (mostly yellow) in Figs 2 and 3. Four of them (Arg 283, Asp 290, Glu 291 and Arg 299) are displayed along a prominent ridge formed by helix 5 and the loop before helix 6. Lys 266 is located just above this ridge on helix 3 (Fig. 3a). Asp 290 and Glu 291 contribute to a patch of negative charge density that also includes Glu 293. An acidic residue is found at position 293 in the FliG proteins from most species, but in *E. coli* and a few other species this residue is uncharged. Its functional importance has not been tested.

In the flagellum, FliG binds to the FliM and FliN proteins to form a 'switch complex' that controls the direction of motor rotation¹². Switching from counterclockwise to clockwise rotation is believed to involve a conformational change in this complex, triggered by the binding of the phosphorylated CheY protein to FliM¹³. Interactions with FliM and FliN are thus likely to be important for holding FliG-C in its proper position and orientation in the motor, and for changing its position or orientation when the motor switches direction. Several observations on mutants indicate that interactions with FliM and

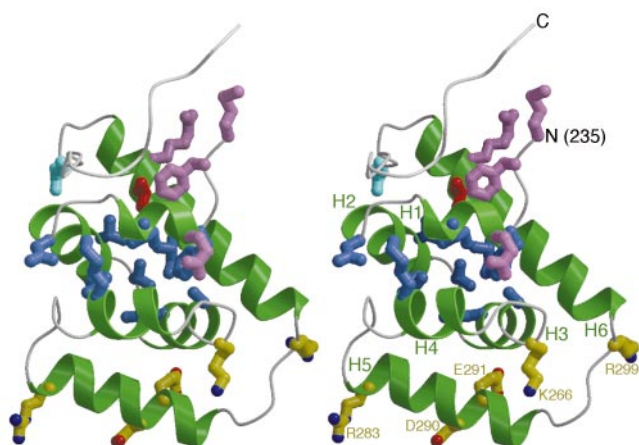


Figure 2 Stereoview ribbon representation of the FliG-C structure. Helices are shown in green and labelled H1-H6. Functionally important charged residues are coloured according to atom type (mostly yellow). Side chains of residues altered in mutants are coloured according to the same scheme as in Fig. 1.

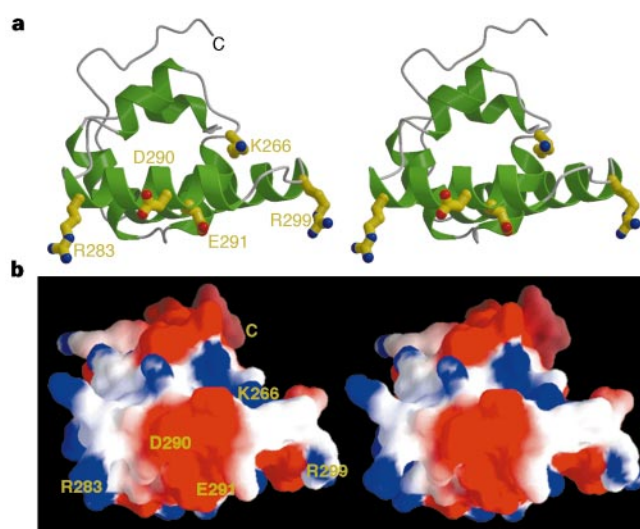


Figure 3 Stereoview of the putative active-site ridge of FliG-C. View direction is from the bottom of Fig. 2. **a**, Ribbon representation showing charged side chains that are important for function and that interact with the stator protein MotA. **b**, Same as **a** but in a GRASP³⁰ representation showing the molecular surface coloured according to electrostatic potential (red, -3kT/e; blue, +3kT/e).

Table 1 Data-collection statistics

	SSRL*(λ 1)	SSRL*(λ 2)	SSRL*(λ 3)	SSRL*(λ 3)	APS†
Wavelength (Å)	0.9800	0.9795	1.0688	0.9252	0.9500
Observed reflections	46,782	52,347	33,292	34,542	56,990
Unique reflections‡	12,546	12,479	9,552	10,185	14,383
$d_{\min}$ (Å)	2.80	2.80	2.50	3.00	2.20
High-resolution shell	2.80–2.85	2.80–2.85	2.50–2.54	3.00–3.05	2.20–2.24
Completeness (%)	99.5 (99.8)	99.6 (99.8)	99.2 (98.3)	99.0 (99.6)	97.1 (92.4)
$R_{\text{sym}}\S$ (%)	7.4 (33.6)	8.9 (45.9)	5.8 (36.2)	9.3 (36.7)	10.1 (41.3)
Average $I/\sigma(I)$	10 (2.5)	9 (2.0)	8.5 (1.5)	9 (2.3)	6 (1.7)
Mosaicity (deg)	0.395	0.385	0.374	0.374	0.537

Data were collected on Quantum-4 CCD detectors, and processed with DENZO and SCALEPACK²⁸. Values in parentheses refer to the highest resolution shell.

* Data collected at the SSRL beamline 1–5.

† Data collected at the APS beamline BIOCARS-14-BM-D.

‡ Number of unique reflections including Bijvoet pairs.

$\S R_{\text{sym}} = 100 \times \sum |I - \langle I \rangle| / \sum I$.

FliN occur near the end of FliG-C opposite the active-site ridge. Mutations in FliG-C that alter clockwise/counterclockwise switching are found in residues Met 235, Phe 236, Glu 239 and Arg 315 (coloured mauve in Fig. 2), which are close to each other and distant from the active-site ridge^{4,6}. A mutation in residue 314 (coloured red in Fig. 2) weakened binding of FliG to FliM in the yeast two-hybrid assay¹⁴. A FliG fragment lacking the active-site ridge and most of the rest of the C-terminal domain (removal of residues 249–335) bound to FliM and FliN in co-precipitation experiments, whereas a smaller fragment lacking some additional residues from the N-terminal region of the ordered FliG-C structure (removal of residues 231–335) did not bind¹⁵. A mutation of residue Glu 322 (coloured turquoise in Fig. 2), isolated as a suppressor of certain mutations in the stator protein MotB, is believed to alter the position of FliG-C relative to the stator¹⁶.

We propose a model for the arrangement of FliG-C units in the flagellar motor based on the observed structure, together with mutational and electron microscopic data (Fig. 4). Electron microscopic studies have shown that FliG molecules in the flagellum form a ring with an outer diameter of ~ 300 Å (refs 17, 18). We propose that the FliG molecules in this ring are arranged with the functionally important charged residues toward the outside, to allow interactions with the MotA/MotB complexes that surround the basal body¹⁹. The active-site ridge is likely to be aligned approximately along the edge of the rotor, so that multiple charged residues of FliG can interact sequentially with the stator as the rotor turns (see below). Given the dimension of the domain in this aspect, simple modelling suggests that a ring of diameter 300 Å could

accommodate as many as 35 copies of FliG, consistent with estimates of FliG copy number in the motor, which range from about 25 to 40 (refs 18, 20, 21).

Interactions between charged residues of FliG and MotA were inferred from patterns of synergism and suppression observed in FliG/MotA double mutants⁵. Certain instances of synergism indicate that all three key basic residues of FliG (Lys 266, Arg 283 and Arg 299), and also Asp 290, interact with Glu 98 of MotA. (Residue 291 of FliG interacts with Arg 90 of MotA.) The residues that interact with Glu 98 of MotA are spaced at some distance from each other on FliG-C, forming a stripe that should roughly define the path traced by the stator as the rotor turns. The key basic residues of FliG-C are not co-linear, however, and seem unable to adopt a linear arrangement even allowing for the flexibility of side chains. We therefore propose that different subsets of charged residues might interact with the stator when the motor turns in the clockwise and counterclockwise directions. This hypothesis is illustrated in Fig. 4c, which shows edge-on views of the rotor with the FliG-C subunits orientated in two ways that bring different subsets of charged residues into alignment along the edge of the rotor. Similar linear arrays of alternating positive and negative charge are present in both orientations, and could provide interactions with MotA that are equivalent during rotation in either direction.

The direction of rotation might be determined by the different surface topographies presented by the two orientations. For example, proton flow through the MotA/MotB channels might drive rotation by forcing conformational changes in the stator that cause

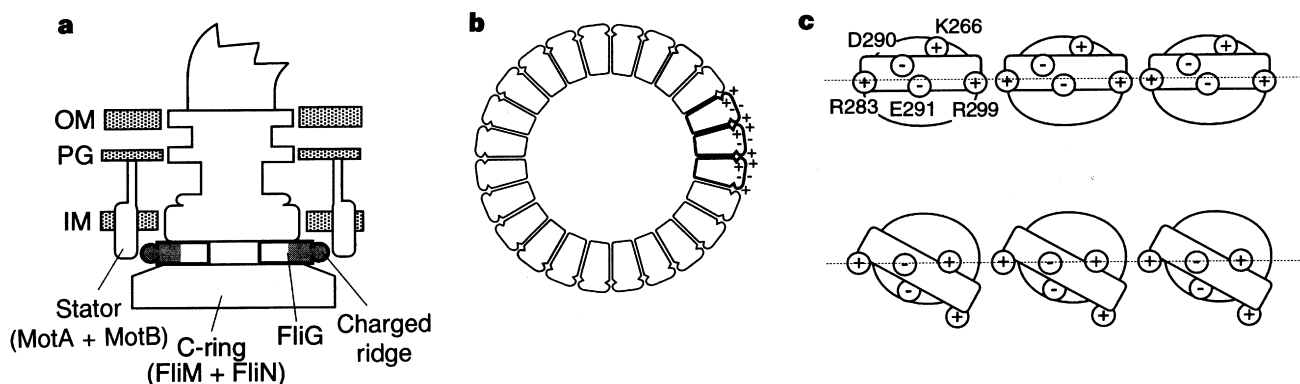


Figure 4 Hypothesis for the location and orientation of FliG-C in the flagellar motor. **a**, Approximate location of FliG relative to other components of the flagellum. OM, outer membrane; PG, peptidoglycan; IM, inner membrane. The C-terminal domain of FliG (FliG-C) is shaded grey and is orientated with the functionally important charged residues towards the outside. Sites of contact with FliM and FliN are inboard from the active-site ridge. **b**, Plan view showing multiple copies of FliG arranged with rotational symmetry and with the active-site ridges aligned roughly parallel to the edge of the rotor. The key charged residues are

indicated with + and – symbols. **c**, Edge-on views of the rotor showing three FliG-C molecules in approximately the same orientation as Fig. 3. Two possible arrangements of FliG-C units are shown that bring different subsets of charged residues into alignment around the edge of the rotor (dotted lines), where they could interact sequentially with a residue(s) of the stator protein MotA. In this model, Arg 283 (and possibly Asp 290) could contact the same residue(s) of MotA in both orientations. Switching between CW and CCW rotation might involve such a change in the orientation of FliG-C.

Table 2 Refinement statistics

Resolution range (Å)	20.0–2.20
Number of protein atoms*	1,523
Number of solvent atoms*	79
R-factor (%)†	23.6
Free R-factor (%)‡	30.4
Bond lengths (Å)§	0.016
Bond angles (deg)§	1.653
(B) (Å ²) main chain	37.39
(B) (Å ²) side chains	40.71
(B) (Å ²) water molecules	44.40
φψ angles (%)§:	
Most favoured	94.6
Additional	5.4

* Non-hydrogen atoms only.

† R-factor = $100 \times \sum (|F_o| - |F_c|) / \sum |F_o|$. All data in the indicated resolution range were used, without application of a cut-off based upon the estimated standard deviation.

‡ Free R-factor is the R-factor for a randomly selected subset (8.0%) of the reflections not included in the refinement calculations.

§ Stereochemistry was assessed with PROCHECK²⁹.

it to press against angled surfaces on the rotor²², with the direction of rotation determined by the topography of the interacting surfaces. Although the exact orientation of FliG-C remains uncertain, it seems clear that many charged residues are present at the edge of the rotor and that this feature is essential for motor rotation. □

Methods

Purification, crystallization and data collection. The crystallized protein consists of residues 209–335 of *T. maritima* FliG with a methionine residue added to the N terminus. This construct was chosen because the corresponding part of the *E. coli* protein was found to be relatively resistant to proteolytic degradation (data not shown). The *T. maritima* FliG gene was amplified by the polymerase chain reaction (PCR) from genomic DNA (a gift from R. Huber, University of Regensburg), using primers designed according to sequence data from The Institute for Genomic Research (www.tigr.org). All of the crystallographic studies were done using selenomethionine-substituted FliG-C, which was overexpressed in the gal-, met-auxotroph B834(DE3) of the BL21 strain of *E. coli*²³. Cells were lysed and FliG-C was purified using thermal denaturation at 80 °C for 10 min to precipitate *E. coli* proteins, followed by ion-exchange (Q-Sepharose, Pharmacia Biotech) and hydrophobic (*n*-octyl-Sepharose, Sigma Chemical) chromatography. N-terminal sequencing of the purified FliG-C indicated that the first seven residues were MDRITTEK, as predicted from the gene sequence.

Purified FliG-C was dialysed in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA and 15 mM 2-mercaptoethanol and concentrated to 8 mg ml⁻¹. The protein was crystallized at 13 °C in sitting drops over a reservoir of 16% isopropanol, 100 mM sodium acetate, pH 5.2, 200 mM CaCl₂, 1 mM EDTA and 10 mM DTT. The drops contained 2 µl of FliG-C solution mixed with 2 µl of the reservoir solution. Data were collected from crystals maintained at 100 K. In preparation for data collection, 50 µl of a cryoprotectant solution (20% isopropanol, 100 mM sodium acetate, pH 5.2, 200 mM CaCl₂, 1 mM EDTA, 10 mM DTT and 20% glycerol) was added to the crystallization drop and crystals were quickly suspended in a small rayon loop and plunged into liquid nitrogen.

The space group is *P*₂₁₂₁₂ with cell dimensions *a* = 50.68 Å, *b* = 62.24 Å and *c* = 88.55 Å. There are two molecules in the asymmetric unit, and the solvent content is ~45%. Data for MAD phase determination were collected at four wavelengths to 3.0 Å resolution from a single crystal at the Stanford Synchrotron Radiation Laboratory. Data from each wavelength were indexed according to the same crystal orientation matrix, and integrated and scaled independently. Scaled data sets from each of the four wavelengths were then scaled together from 20.0 to 3.0 Å resolution. A different crystal gave higher-resolution (2.2 Å) diffraction at the Advanced Photon Source synchrotron, where data were collected at a single wavelength for refinement.

Structure determination and refinement. Six ordered selenomethionine positions, of a total of 12 methionine residues in the asymmetric unit, were determined from difference Patterson and Fourier maps using XtalView²⁴. Selenium positions were refined in PHASES²⁵, and the mean figure of merit was 0.68 to 2.8 Å resolution. Solvent flattening was done with PHASES to a mean

figure of merit of 0.89, greatly aiding interpretation of the electron density map. Rounds of refinement with X-PLOR²⁶ were interspersed with manual model building using the program O²⁷. The final refinement cycles, against all data from 20.0 to 2.2 Å resolution, included a bulk-solvent correction, overall anisotropic *B*-factor correction, and non-crystallographic symmetry (n.c.s.) restraints on 86 Cα atoms. The current model includes residues 235–335 of molecule 1, residues 235–323 of molecule 2, and 79 water molecules. The R-factor is 23.6% (*R*_{free} is 30.4%), with good stereochemistry (Table 2). Residues with ill-defined electron density (residues 208–234 of both molecules and 324–335 of molecules 2) were not included in the model.

Received 12 April; accepted 13 June 1999.

- Blair, D. F. How bacteria sense and swim. *Annu. Rev. Microbiol.* **49**, 489–522 (1995).
- Macnab, R. M. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (eds Neidhardt, F. C.) 123–146 (American Society for Microbiology, Washington DC, 1996).
- DeRosier, D. J. The turn of the screw: the bacterial flagellar motor. *Cell* **93**, 17–20 (1998).
- Lloyd, S. A. & Blair, D. F. Charged residues of the rotor protein FliG essential for torque generation in the flagellar motor of *Escherichia coli*. *Mol. Biol.* **266**, 733–744 (1997).
- Zhou, J., Lloyd, S. A. & Blair, D. F. Electrostatic interactions between rotor and stator in the bacterial flagellar motor. *Proc. Natl. Acad. Sci. USA* **95**, 6436–6441 (1998).
- Irikura, V. M., Kihara, M., Yamaguchi, S., Sockett, H. & Macnab, R. M. *Salmonella typhimurium* fliG and fliN mutations causing defects in assembly, rotation, and switching of the flagellar motor. *J. Bacteriol.* **175**, 802–810 (1993).
- Huber, R. *et al.* *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90 °C. *Arch. Microbiol.* **144**, 324–333 (1986).
- Yamaguchi, S., Fujita, H., Ishihara, A., Aizawa, S. & Macnab, R. M. Subdivision of flagellar genes of *Salmonella typhimurium* into regions responsible for assembly, rotation, and switching. *J. Bacteriol.* **166**, 187–193 (1986).
- Hutchinson, E. G. & Thornton, J. M. PROMOTIF—a program to identify and analyze structural motifs in proteins. *Prot. Sci.* **5**, 212–220 (1996).
- Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
- Zhou, J. & Blair, D. F. Residues of the cytoplasmic domain of MotA essential for torque generation in the bacterial flagellar motor. *J. Mol. Biol.* **273**, 428–439 (1997).
- Yamaguchi, S. *et al.* Genetic evidence for a switching and energy-transducing complex in the flagellar motor of *Salmonella typhimurium*. *J. Bacteriol.* **168**, 1172–1179 (1986).
- Welch, M., Oosawa, K., Aizawa, S.-I. & Eisenbach, M. Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. *Proc. Natl. Acad. Sci. USA* **90**, 8787–8791 (1993).
- Marykwas, D. L. & Berg, H. C. A mutational analysis of the interaction between FliG and FliM, two components of the flagellar motor of *Escherichia coli*. *J. Bacteriol.* **178**, 1289–1294 (1996).
- Tang, H., Braun, T. F. & Blair, D. F. Motility protein complexes in the bacterial flagellar motor. *J. Mol. Biol.* **261**, 209–221 (1996).
- Garza, A. G., Biran, R., Wohlschlegel, J. & Manson, M. D. Mutations in *motB* suppressible by changes in stator or rotor components of the bacterial flagellar motor. *J. Mol. Biol.* **258**, 270–285 (1996).
- Francis, N. R., Sosinsky, G. E., Thomas, D. & DeRosier, D. J. Isolation, characterization and structure of bacterial flagellar motors containing the switch complex. *J. Mol. Biol.* **235**, 1261–1270 (1994).
- Zhao, R., Pathak, N., Jaffe, H., Reese, T. S. & Khan, S. FliN is a major structural protein of the C-ring in the *Salmonella typhimurium* flagellar basal body. *J. Mol. Biol.* **261**, 195–208 (1996).
- Khan, S., Dapice, M. & Reese, T. S. Effects of *mot* gene expression on the structure of the flagellar motor. *J. Mol. Biol.* **202**, 575–584 (1988).
- Jones, C. J., Macnab, R. M., Okino, H. & Aizawa, S.-I. Stoichiometric analysis of the flagellar hook-(basal-body) complex of *Salmonella typhimurium*. *J. Mol. Biol.* **212**, 377–387 (1990).
- Francis, N. R., Irikura, V. M., Yamaguchi, S., DeRosier, D. J. & Macnab, R. M. Localization of the *Salmonella typhimurium* flagellar switch protein FliG to the cytoplasmic M-ring face of the basal body. *Proc. Natl. Acad. Sci. USA* **89**, 6304–6308 (1992).
- Braun, T. F. *et al.* Function of proline residues of MotA in torque generation by the flagellar motor of *Escherichia coli*. *J. Bacteriol.* **181**, 3542–3551 (1999).
- Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. The crystal structure of the globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362**, 219–223 (1993).
- McRee, D. E. A visual protein crystallographic software system for X11/XView. *J. Molec. Graph.* **10**, 44–46 (1992).
- Furey, W. & Swaminathan, S. PHASES: A program package for the processing and analysis of diffraction data from macromolecules. *Methods Enzymol.* **277**, 590–620 (1997).
- Brünger, A. T. X-PLOR Version 3.843, a system for X-ray Crystallography and NMR (Yale University, New Haven, 1996).
- Jones, T. A., Zhou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and locations of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **176**, 307–326 (1997).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Acknowledgements. We thank H. Bellamy, A. Kolatkar, M. Mathews, J. R. Knowlton and J. D. Phillips for assistance with data collection, and R. Huber for *T. maritima* genomic DNA. Preliminary sequence data were obtained from The Institute for Genomic Research website at <http://www.tigr.org>. Sequencing of *T. maritima* was accomplished with support from the US Department of Energy (DoE). The SSRL Biotechnology Program is supported by the US National Institutes of Health (NIH), National Center for Research Resources, Biomedical Technology Program, and by the DoE Office of Biological and Environmental Research. Use of APS was supported by the DoE Basic Energy Sciences, Office of Science. Use of the BioCARS Sector 14 was supported by the NIH National Center for Research Resources. S.A.L. was partially supported by an NIH training grant. Supported by grants from the NSF (to D.F.B.) and NIH (to C.P.H.).

Correspondence and requests for materials should be addressed to D.F.B. (e-mail: blair@bioscience.utah.edu) or C.P.H. (e-mail: chris@snowbird.med.utah.edu). Coordinates and diffraction data have been deposited in the Brookhaven Protein Bank under accession code 1QC7.

Structure of cytochrome *c* nitrite reductase

Oliver Einsle^{*,†}, Albrecht Messerschmidt^{*}, Petra Stach[†],
Gleb P. Bourenkov[‡], Hans D. Bartunik[‡], Robert Huber^{*}
& Peter M. H. Kroneck[†]

^{*} Max-Planck-Institut für Biochemie, Abteilung Strukturforschung,
Am Klopferspitz 18a, 82152 Martinsried, Germany

[‡] MPG-ASMB c/o DESY, Arbeitsgruppe Proteindynamik, Notkestrasse 85,
22603 Hamburg, Germany

[†] Universität Konstanz, Fakultät für Biologie, 78457 Konstanz, Germany

The enzyme cytochrome *c* nitrite reductase catalyses the six-electron reduction of nitrite to ammonia as one of the key steps in the biological nitrogen cycle¹, where it participates in the anaerobic energy metabolism of dissimilatory nitrate ammonification². Here we report on the crystal structure of this enzyme from the microorganism *Sulfurospirillum deleyianum*, which we solved by multiwavelength anomalous dispersion methods. We propose a reaction scheme for the transformation of nitrite based on structural and spectroscopic information. Cytochrome *c* nitrite reductase is a functional dimer, with 10 close-packed haem groups of type *c* and an unusual lysine-coordinated high-spin haem at the active site. By comparing the haem arrangement of this nitrite reductase with that of other multi-haem cytochromes, we have been able to identify a family of proteins in which the orientation of haem groups is conserved whereas structure and function are not.

The nitrogen cycle has received much attention in recent years because of its ecological importance. Here, N_xO_y compounds serve as electron acceptors of anaerobic respiratory chains in bacteria that take advantage of the favourable redox potentials ($E^{\circ'}$), such as $E^{\circ'}(NO_3^-/NO_2^-) = +0.43$ V, $E^{\circ'}(NO_2^-/NO) = +0.35$ V or $E^{\circ'}(NO_2^-/NH_4^+) = +0.34$ V for energy conservation^{3,4}. In nitrate ammonification, nitrate is reduced to nitrite as the only liberated intermediate. Nitrite is subsequently converted to ammonia in a six-electron step by cytochrome *c* nitrite reductase⁵.

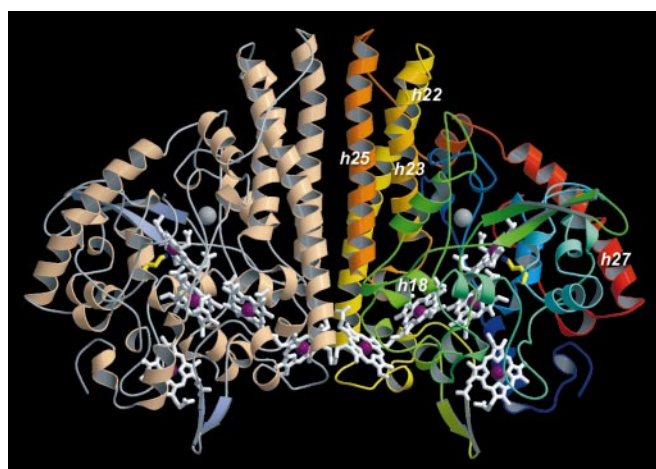
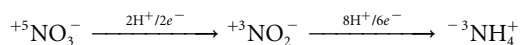


Figure 1 The nitrite reductase dimer. A front view with the dimer axis orientated vertically, the five haems in each monomer (white), the Ca^{2+} ions (grey) and Lys 133 which coordinates the active-site iron atom (yellow). In the right monomer, the protein chain is coloured blue from the amino-terminal end to red at the carboxy-terminal end, in the left monomer according to secondary structure. The dimer interface is dominated by three long α -helices per monomer. All haems in the dimer are covalently attached to the protein and their iron atoms are arranged almost in a plane parallel to the plane of the paper.



These multihaem enzymes have been isolated from various microorganisms, including *Desulfovibrio desulfuricans*⁶, *Wolinella succinogenes*⁷⁻⁹ and *S. deleyianum*^{10,11}, which form ammonia from nitrate in a dissimilatory process. A comparison of molecular properties indicates that the respiratory ammonia-forming nitrite reductases may constitute a group of homologous enzymes¹². Darwin *et al.*¹³ reported on the sequence of the structural gene of the *Escherichia coli* cytochrome *c* nitrite reductase, which carries four Cys-X₁-X₂-Cys-His haem-binding motifs. In that sequence, they observed a Cys-X₁-X₂-Cys-Lys motif which originally was not assigned to a haem-binding site, in agreement with the results of the haem and iron determination for nitrite reductase from *S. deleyianum* and *W. succinogenes*¹⁰. Most recently, it was established that a haem is bound covalently to the cysteine-lysine motif in the *E. coli* enzyme^{1,14}.

S. deleyianum cytochrome *c* nitrite reductase has a high specific activity, with up to 1,050 μ mol NO_2^- per min per mg (pH 7); this has led to the development of an amperometric biosensor^{15,16}. The enzyme from *S. deleyianum* also reduced sulphite to hydrogen sulphide, again a six-electron transfer process¹⁷. The crystal structure shows that the enzyme is a homodimer of $\sim 100 \text{ \AA} \times 80 \text{ \AA} \times 50 \text{ \AA}$, with 10 haemes in remarkably close packing (Fig. 1). Three monomers form the asymmetric unit of the crystal; two of these form a dimer through a non-crystallographic axis and the third uses a crystal axis for dimer formation. The protein folds into one compact domain, with α -helices as the predominant secondary structural motif, ranging from short helical turns to four long helices at the carboxy-terminal end of the peptide chain (Fig. 1). There are eight 3_{10} helices three or four residues long, seven of which occur within the first 200 residues. β -sheet structures are found only in two short, antiparallel strands, where one is part of a funnel-like cavity leading to the active site. Dimer formation is mediated by helices *h22* and *h25*, with helix *h25* as the key element, as it interacts with its counterpart in the other monomer over its full length of 28 residues, corresponding to 42 $\text{\AA}$. The contact surface has an area of 1,674 $\text{\AA}^2$, or 8.5% of the total accessible surface area of a monomer.

High-resolution structures are available for the copper nitrite reductase¹⁸ and cytochrome *cd*₁ nitrite reductase¹⁹ from denitrifying bacteria. Both enzymes catalyse the reduction of nitrite to nitric oxide, and their structures show no similarities whatsoever either to each other or to cytochrome *c* nitrite reductase. Apparently, the use of nitrite in bacterial energy conservation has evolved in three independent pathways.

The five haems in the monomer of cytochrome *c* nitrite reductase are in close contact, with Fe-Fe distances of between 9 and 12.8 $\text{\AA}$ (Fig. 2). They are arranged as a group of three, almost coplanar,

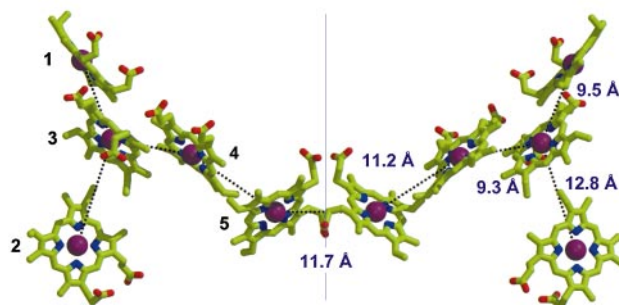


Figure 2 Haem arrangement. The overall orientation corresponds to Fig. 1, with the active site located at haem 1 and the line indicating the dimer interface. Haems in the left monomer are numbered according to their attachment to the protein chain. In the right monomer, the Fe-Fe distances ($\text{\AA}$) between the haems are given. Haems 5 interact across the dimer interface with a distance closer than haems 2 and 3 within each monomer.

haems (1, 3, 4), with haem 1 forming the active site. Haems 2 and 5 are farther apart and are not coplanar with haems 1, 3 and 4. All haems except haem 1 are *bis*-histidinyl-coordinated and are linked to the peptide backbone by thioether bonds to the cysteine residues of a classical haem-binding motif for periplasmic proteins, Cys-X₁-X₂-Cys-His. With edge-to-edge distances below 4 Å, haems 1, 3 and 4 are close enough to allow direct π -electron interaction of the porphyrin rings. The propionate side chains of haem 1 form part of the active-site cavity, while those of haem 4 are exposed to the solvent and those of haem 3 are hydrogen-bonded inside the protein. All porphyrins show a slight distortion from ideal planarity, most strongly in haem 2 and least in the active-site haem 1. Haem 2 could function as the entry point for electrons. The short Fe-Fe distance of 11.7 Å between haems 5 within the dimer suggests an electronic interaction of these centres which might be functionally relevant¹¹. Furthermore, both haems interact directly through hydrogen bonds between their propionate side chains.

The site of nitrite reduction is clearly haem 1, with the N ζ atom of Lys 133, replacing a histidine in the classical binding motif (Fig. 3a), and an oxygen atom of a sulphate anion. This type of haem-iron coordination is observed here for the first time, to our knowledge. An additional electron-density maximum detected close to the active site was assigned to Ca²⁺. This was confirmed by inductively coupled plasma atomic emission spectroscopy, which yielded 1.0 ± 0.1 atoms of Ca²⁺ per nitrite reductase monomer. The Ca²⁺ ion is coordinated by the carboxy group of Glu 216 in a bidentate manner and by Gln 281, two peptide carbonyl oxygens and two water molecules. Sulphate binds to the iron with an oxygen atom at a bond length of 2.0 Å. It is further coordinated to both His 282 and Tyr 217 with a single oxygen atom and to a water molecule which is in turn coordinated to the two propionate side chains of haem 1. Arginine 113 also forms a hydrogen bond to one of the propionates and is bonded to the sulphate by another water molecule. The roof of the active-site cavity is formed by Phe 91, Lys 279, Tyr 95, Ala 404

and Gln 281, which coordinates the Ca²⁺ by its carboxamide oxygen such that the amino group faces the active site. This may contribute to the positive electrostatic surface potential surrounding the active site (Fig. 4).

S. deleyianum cytochrome *c* nitrite reductase reduces not only nitrite to ammonia, but also the potential intermediates NO

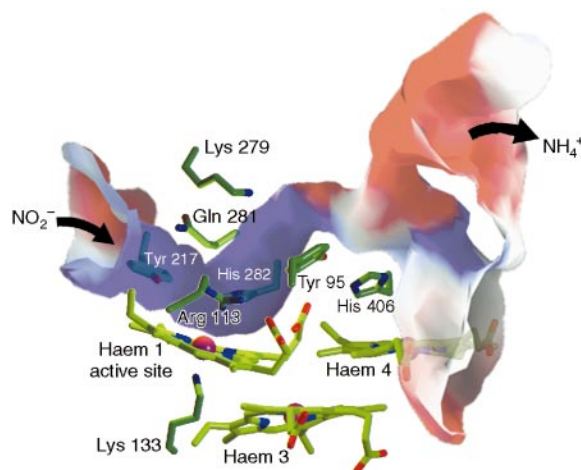


Figure 4 The active-site channel. Apart from the channel leading from the protein surface to the catalytic site, a second channel has been detected that reaches the protein surface on the opposite side of the molecule. The whole channel is rendered as a semi-transparent surface and coloured according to the electrostatic surface potential; blue for a positive and red for a negative potential (cutoff +10 and -10, respectively). In contrast to the immediate surroundings of the active-site haem, the surface potential is drastically changed in the presumed exit channel for ammonia.

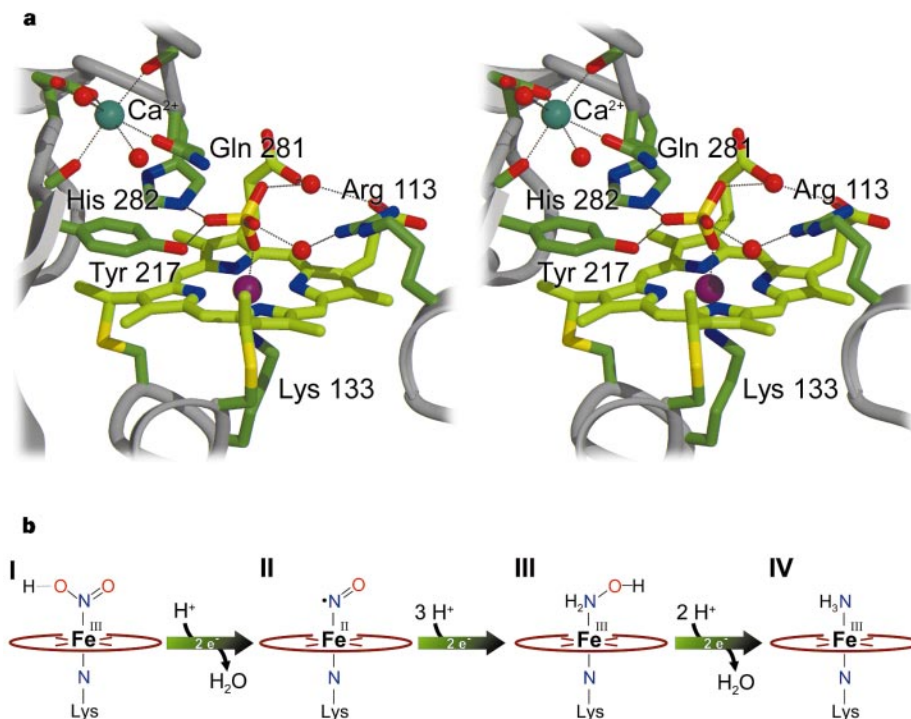


Figure 3 The active site of the enzyme. **a**, A view from the protein surface down the active-site channel. Tyr 217 and His 282 bind sulphate, an inhibitor of enzyme activity. A Ca²⁺ ion close by bridges stretches of protein that hold these two residues. **b**, A minimal reaction scheme. Structure **I** represents the catalytic haem

1 in the oxidized state, with nitrite bound. In structure **II** the iron is reduced and NO is bound. Structure **III** shows the putative hydroxylamine intermediate, and in the final-step ammonia is formed (structure **IV**) and then released as an ammonium cation.

and hydroxylamine. No intermediates are released during nitrite turnover. Obviously, the active site accommodates anions and uncharged molecules and releases the ammonium cation only after the full six-electron reduction. The preference for anions is reflected by a positive electrostatic potential around and inside the active-site cavity. This is induced by the residues forming the cavity: Tyr 127, His 282, Arg 113, Gln 281 and Lys 279. These residues will serve as stores for protons required for the reduction of nitrite to ammonia and can be resupplied by water molecules. Considering the good accessibility of the active site for water molecules and the presumably lowered pH on the periplasmic side of the cytoplasmic membrane, the product of nitrite reduction will be the positively charged ammonium ion rather than uncharged ammonia. The cationic product might make use of a second channel leading to the protein surface opposite to the entry channel. This second channel is lined by His 406 and Tyr 95 and filled with coordinated water molecules (Fig. 4). It branches before reaching the protein surface and ends with both arms in areas possessing a significantly negative electrostatic surface potential. The existence of separate pathways for substrate and product with matched electrostatic potential could also contribute to the enzyme's high specific activity compared with the sirohaem-containing nitrite reductases that catalyse the same reaction.

Cytochrome *c* nitrite reductase is believed to form a membrane-associated complex with a further multihaem cytochrome that functions as a direct electron donor²⁰. Such a complex would allow fast electron transfer, not limited by diffusion and docking of redox partners.

Although the structure of the enzyme with bound substrate is not yet available, analogies with other types of nitrite and sulphite

reductases and the arrangement of active-site residues suggest that nitrite and sulphite bind to the iron with their nitrogen or sulphur atoms, respectively. One oxygen atom will be bound by His 282 and Tyr 217, in analogy to cytochrome *cd*₁ nitrite reductase, which shows a 2-fold coordination by two histidine residues²¹. The Fe^{II}-NO adduct (Fig. 3b; structure II) has been identified under turnover conditions by rapid-freeze electron paramagnetic resonance spectroscopy. It shows an intense resonance at $g = 2.014$, and a hyperfine splitting $A^N = 1.55$ mT. In contrast to cytochrome *cd*₁, NO is not released from cytochrome *c* nitrite reductase, but is further reduced to ammonia.

Although there is no significant sequence homology, the arrangement of the haems in cytochrome *c* nitrite reductase is very similar to that of hydroxylamine oxidoreductase (HAO) from *Nitrosomonas europaea*, a trimeric octahaem cytochrome that catalyses the oxidation of hydroxylamine to nitrite²². The five haems of nitrite reductase align to haems 4–8 of HAO with a root-mean-square deviation of 2.00 Å, and the three dimer-forming helices *h*22, *h*23 and *h*25 of nitrite reductase are topologically equivalent to helices α 20, α 22/ α 23 and α 24 of HAO (Fig. 5). Furthermore, helix *h*18 of nitrite reductase, which runs along the three coplanar haems, corresponds to helix α 15 of HAO. These similarities strongly indicate a distant evolutionary relatedness for the two proteins.

Whereas a classical protoporphyrin IX haem containing a novel lysine coordination forms the active site of the nitrite reductase, HAO features the P460-type haem with a covalent link between a haem *meso* CHC and C ϵ 1 of a tyrosine residue of an adjacent chain in the functional trimer. The fifth coordination position of P460 is occupied by a histidine. As HAO functions as an oxidase and nitrite reductase as a reductase, the replacement of histidine by lysine can

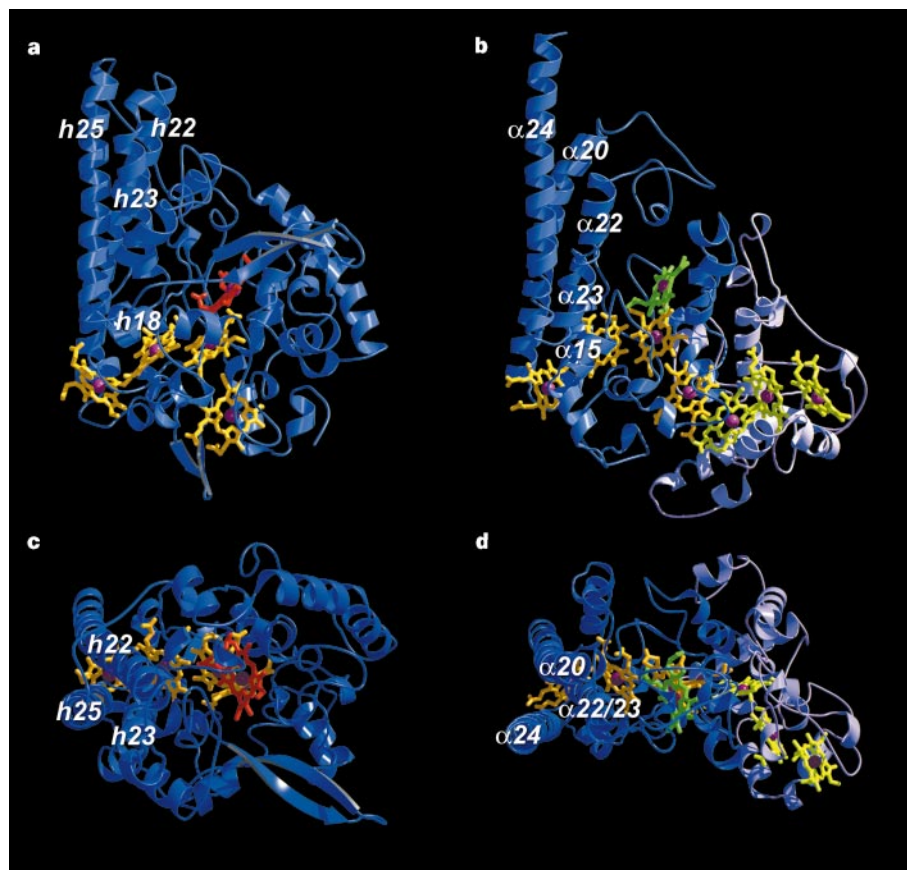


Figure 5 Comparison between nitrite reductase (NiR) and hydroxylamine oxidoreductase (HAO). **a**, Front view of NiR. **b**, Front view of HAO. **c**, Top view of NiR. **d**, Top view of HAO. Corresponding haems are drawn in orange, additional haems in HAO in yellow. The red haem in NiR is the lysine-coordinated haem 1, the

green haem in HAO is haem P460. The helices indicated are the only elements of secondary structure that can be sensibly superimposed. The part of the protein chain of HAO that binds the first three haems and has no counterpart in NiR is drawn in violet.

Table 1 MAD phasing statistics

Wavelength (Å)	Resolution limits (Å)	Completeness (%)	R_{merge} (%) [*]	f'	f''	Phasing power _{iso} [†]	Phasing power _{ano} [‡]	R_{Cullis} [§]	R_{Kraut}
1.0500	1.9	98.9	10.5	0.2	1.7		0.95		21.9
1.7362	2.3	99.9	11.8	−5.8	6.0	2.71	0.65	59.6	66.7
1.7390	2.3	98.5	10.0	−10.0	3.5	3.00	1.12	55.8	52.0

^{*} $R_{\text{merge}} = \sum_{hkl} |I| - I / \sum_{hkl} |I|$.

[†]Isomorphous phasing power = $\sum_{hkl} |F_H| / \sum_{hkl} |E|$, where $|F_H|$ is the calculated structure-factor amplitude of the heavy-atom structure and E is the residual lack of closure error.

[‡]Anomalous phasing power = $2 \sum_{hkl} |F_H^2| / \sum_{hkl} |E|$ where $|F_H^2|$ is the anomalous-contribution amplitude.

[§] $R_{\text{Cullis}} = \sum_{hkl} |F_{PH} \pm F_P| - F_{H(\text{calc})} / \sum_{hkl} |F_{PH} \pm F_P|$ for centric reflections and isomorphous contributions.

^{||} $R_{\text{Kraut}} = \sum_{hkl} |F_{PH}| - |F_P^2 + F_{H(\text{calc})}^2| / \sum_{hkl} |F_{PH}|$ for anomalous scattering contributions.

be expected to raise the redox potential of the haem, generating an electron sink in the system of redox-active cofactors. Furthermore, with a coordinated lysine, the product complex containing two sp^3 -hybridized nitrogen ligands at haem 1 might increase the driving force of the reaction.

The overall set-up of the two proteins shows remarkable differences. Whereas HAO is a functional trimer, nitrite reductase acts as a dimer. The trimerization of HAO leads to the formation of a ring of 18 haems, 6 from each monomer²². The P460-haem is situated above this ring and is close to haem 8 of the adjacent monomer. Haem 1 is not involved in the formation of the ring. In nitrite reductase, haem 5, the topological equivalent of haem 8 in HAO, is the one that interacts over the dimer interface. In both cases the result is the ability to distribute electrons quickly over a larger number of haem centres.

The only feature that is strictly conserved between nitrite reductase and HAO, apart from the arrangement of the haems, is a histidine at the active site at the beginning of a topologically conserved helix: His 282 at the beginning of helix *h*18 in nitrite reductase and His 268 in helix α 15 in HAO. In nitrite reductase, this histidine binds the sulphate anion, and its conservation in both proteins suggests a role in the reduction of nitrite to hydroxylamine, or at least in the removal of the first oxygen atom in the form of water from nitrite in the reaction sequence of the reductase (Fig. 3b). Nitrite reductase catalyses a six-electron reduction, whereas HAO performs a four-electron oxidation. The two-electron step from hydroxylamine to ammonia occurs only in the reductase. The features where the two active sites differ significantly are probably involved in this part of the reaction. The main differences are obviously Tyr 217 of the reductase, which binds to the sulphate ion, the coordination of Lys 133 to haem 1 and the orientation of its propionate side chains. Conversion of hydroxylamine to ammonia requires the removal of an oxygen atom, leaving a bound $\text{NH}_2^{\cdot}$ radical, which would then receive an electron plus a proton from Tyr 217 to yield ammonia and a tyrosyl radical which must persist until a further electron arrives. In HAO the covalently linked Tyr 467 comes closer to the haem iron by means of its oxygen atom (3.7 Å) than does Tyr 217 in the sulphate-bound form of nitrite reductase (5.2 Å). However, in HAO the oxygen points away from the iron and

forms a rather unfavourable angle for electronic interaction with any iron ligand, whereas in the reductase a simple rotation around the tyrosine's χ 1-angle would bring it much closer to the iron.

Note that the structure of the tetrahaem cytochrome c_{554} , the redox partner of HAO²³, also shows a haem arrangement similar to that of HAO. As in nitrite reductase, only the four haems of cytochrome c_{554} and their immediate protein surroundings align with haems 3–6 of HAO while the rest of the structure of the two proteins differs completely. Although cytochrome c_{554} has no known catalytic activity, it shows only 5-fold coordination of the iron atom in the haem group that corresponds to the P460 haem of HAO and consequently also to the active-site haem of the reductase. The special arrangement of haem centres in this emerging group of multihaem proteins seems to be of functional significance, and represents a motif whose exact function remains unknown. □

Methods

Protein crystallization and data collection. The protein was purified as described^{10,11}. Small crystal plates of cytochrome *c* nitrite reductase (NiR) were grown by sitting-drop vapour diffusion from 40 mg ml^{−1} protein solution, 2% PEG 6000 and 0.1 M HEPES/NaOH at pH 7.5. The plates were used for microseeding into pre-equilibrated drops of 5 mg ml^{−1} of protein solution and 10% PEG 4000, 0.13 M ammonium sulphate and 0.1 M MES/NaOH, pH 7.5. These crystals belonged to spacegroup $P2_12_12_1$, with unit cell constants of $a = 107.49$ Å, $b = 185.34$ Å and $c = 92.19$ Å. With three monomers of 58.6 kilodaltons per asymmetric unit, this results in a solvent content of 53%. A crystal of $150 \times 100 \times 50 \mu\text{m}^3$ cooled to 100 K in a nitrogen stream was sufficient to collect data at three wavelengths, which eventually allowed us to solve the structure to a final resolution of 1.9 Å. To avoid crystal damage by the nitrogen stream cooling, we added 20% methylpentanediol to the crystals in their mother liquor. This was done in steps of 5% with 15 min equilibration time. Successfully cryo-cooled crystals showed no significant decay in the X-ray beam. Data were collected from a single crystal on beamline BW6 at the German Electron Synchrotron (DESY) in Hamburg, Germany, on a Mar Research CCD detector at three wavelengths: $\lambda_1 = 1.050$ Å as a remote dataset and $\lambda_2 = 1.736$ Å and $\lambda_3 = 1.739$ Å to maximize the anomalous f'' and the dispersive f' contribution of the K-shell absorption edge of iron, respectively. Data sets were integrated, scaled and merged using the *hkl-suite*²⁴.

Gene sequence determination. The structural gene encoding cytochrome *c* nitrite reductase, called *nrfA* following the nomenclature established for *E. coli*¹³, was identified using the polymerase chain reaction (PCR). Heterologous oligonucleotides, derived from sequences of peptide fragments of purified NiR, were used as PCR primers to produce a first fragment of the gene. This fragment was labelled with digoxigenin and used as a probe to screen a genomic library constructed by partial digestion of genomic DNA of *S. deleyianum* with the restriction endonuclease *Sau*3AI. In this screening, a clone was identified which contained most of the gene. The missing part was then found by inverse PCR.

Multiwavelength anomalous dispersion (MAD) phasing, model building and refinement. Anomalous difference Patterson maps were calculated from the data collected at the f'' maximum of the iron-absorption edge. Correlated peaks that appeared in all three Harker sections were chosen and used for phasing with MLPHARE²⁵. The highest peaks from the resulting Fourier map were analysed for consistency with the anomalous difference Patterson map and, if correct, used for a new cycle of phase calculations. Thus, it was possible successively to find 15 iron positions which divided into three groups of five

Table 2 Refinement statistics

No. of independent reflections [*]	143,650
Resolution range	25.0–1.90 Å
No. of non-hydrogen atoms	13,431
Protein residues	1,419
Disordered side chains	26
Water molecules	1,485
Sulphate molecules	9
$I/\sigma(I)$ overall (2.0–1.9 Å)	16.6 (2.5)
Crystallographic <i>R</i> -factor	0.184
Free <i>R</i> -factor†	0.220
Average <i>B</i> -factor for protein	25.1 Å ²
Average <i>B</i> -factor for waters	35.9 Å ²
r.m.s.d. in bond lengths	0.011 Å
r.m.s.d. in bond angles	1.415°

^{*}In the dataset used for refinement, which was collected at $\lambda = 1.0500$ Å, Friedel pairs were merged for refinement.

[†]The free *R*-factor was calculated from 5% of the data, removed at random before the refinement was carried out.

and produced a sensible packing when the spacegroup's symmetry operations were applied.

When trying to obtain phases for model building, MLPHARE did not produce interpretable electron-density maps, so the program SHARP²⁶ was used, resulting in a phasing power of 3.0 and a figure of merit of 0.54 over the whole resolution range of the anomalous data (Table 1). The according values from MLPHARE were 1.6 and 0.42, respectively. Solvent flattening was carried out with the program SOLOMON²⁷.

For crystallographic positional refinement and density-map calculations, X-PLOR was used²⁸ with Engh and Huber parameters²⁹. In addition, torsion-angle refinement and molecular dynamics, both based on maximum-likelihood algorithms, were carried out with the developmental version 0.3 of the crystallography and NMR system CNS³⁰. After building the complete model, the structure refined to a crystallographic *R*-factor (*R*_{free}) of 0.281 (0.328). 1,485 water molecules and 9 sulphate ions were built and the *R*-factor (*R*_{free}) was refined to its final value of 0.184 (0.220) at a resolution of 1.9 Å (Table 2). Noncrystallographic symmetry restraints were applied only in the first refinement cycles.

Received 16 April; accepted 27 May 1999.

- Berks, B. C., Ferguson, S. J., Moir, J. W. B. & Richardson, D. J. Enzymes and associated electron transport systems that catalyse the respiratory reduction of nitrogen oxides and oxyanions. *Biochim. Biophys. Acta—Bioenergetics* **1232**, 97–173 (1995).
- Cole, J. A. & Brown, C. M. Nitrite reduction to ammonia by fermentative bacteria: a short circuit in the biological nitrogen cycle. *FEMS Microbiol. Lett.* **7**, 65–72 (1980).
- Zumft, W. G. Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533 (1997).
- Thauer, R. K., Jungermann, K. & Decker, K. Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol. Rev.* **41**, 100–180 (1977).
- Cole, J. A. Nitrate reduction to ammonia by enteric bacteria: redundancy, or a strategy for survival during oxygen starvation? *FEMS Microbiol. Lett.* **136**, 1–11 (1996).
- Liu, M. C. & Peck, H. D. J. The isolation of a hexaheme cytochrome from *Desulfovibrio desulfuricans* and its identification as a new type of nitrite reductase. *J. Biol. Chem.* **256**, 13159–13164 (1981).
- Liu, M.-C., Liu, M.-Y., Payne, W. J., Peck, H. D. Jr & LeGall, J. Wolinella succinogenes nitrite reductase: purification and properties. *FEMS Microbiol. Lett.* **19**, 201–206 (1983).
- Schroeder, I. et al. The membranous nitrite reductase involved in the electron transport of Wolinella succinogenes. *Arch. Microbiol.* **140**, 380–386 (1985).
- Blackmore, R., Robertson, A. M. & Brittain, T. The purification and some equilibrium properties of the nitrite reductase of the bacterium Wolinella succinogenes. *Biochem. J.* **233**, 547–552 (1986).
- Schumacher, W. & Kroneck, P. M. H. Dissimilatory hexaheme c nitrite reductase of "Spirillum" strain 5175: purification and properties. *Arch. Microbiol.* **156**, 70–74 (1991).
- Schumacher, W., Hole, U. & Kroneck, P. M. H. Ammonia-forming cytochrome c nitrite reductase from *Sulfurospirillum deleyianum* is a tetraheme protein: new aspects of the molecular composition and spectroscopic properties. *Biochem. Biophys. Res. Commun.* **205**, 911–916 (1994).
- Schumacher, W., Neese, F., Hole, U. & Kroneck, P. M. H. in *Transition Metals in Microbial Metabolism* (eds Winkler, G. & Carrano, C. J.) 329–356 (Harwood Academic, Amsterdam, 1997).
- Darwin, A. et al. Regulation and sequence of the structural gene for cytochrome c₅₅₂ from *Escherichia coli*: not a hexaheme but a 50 kDa tetraheme nitrite reductase. *Mol. Microbiol.* **9**, 1255–1265 (1993).
- Eaves, D. J. et al. Involvement of products of the *nrEF* genes in the covalent attachment of haem c to a novel cysteine–lysine motif in the cytochrome c₅₅₂ nitrite reductase from *Escherichia coli*. *Mol. Microbiol.* **28**, 205–216 (1998).
- Moreno, C. et al. Electrochemical studies of the hexaheme nitrite reductase from *Desulfovibrio desulfuricans* ATCC 27774. *Eur. J. Biochem.* **212**, 79–86 (1993).
- Strehlitz, B. et al. A nitrite sensor based on a highly sensitive nitrite reductase mediator-coupled amperometric detection. *Anal. Chem.* **68**, 807–816 (1996).
- Einsle, O., Schumacher, W., Kurun, E., Nath, U. & Kroneck, P. M. H. in *Biological Electron Transfer Chains: Genetics, Composition and Mode of Operation* (eds Canters, G. W. & Vijgenboom, E.) 197–208 (Kluwer Academic, Dordrecht, 1998).
- Godden, J. W. et al. The 2.3 Å X-ray structure of nitrite reductase from *Achromobacter cycloclastes*. *Science* **253**, 438–442 (1991).
- Fülöp, V., Moir, J. W. B., Ferguson, S. J. & Hajdu, J. The anatomy of a bifunctional enzyme: structural basis for reduction of oxygen to water and synthesis of nitric oxide by cytochrome cd₁. *Cell* **81**, 369–377 (1995).
- Hussain, H., Grove, J., Griffiths, L., Busby, S. & Cole, J. A seven-gene operon essential for formate-dependent nitrite reduction to ammonia by enteric bacteria. *Mol. Microbiol.* **12**, 153–163 (1994).
- Williams, P. A. et al. Haem–ligand switching during catalysis in crystals of a nitrogen-cycle enzyme. *Nature* **389**, 406–412 (1997).
- Igarashi, N., Moriyama, H., Fujiwara, T., Fukumori, Y. & Tanaka, N. The 2.8 Å structure of hydroxylamine oxidoreductase from a nitrifying chemoautotrophic bacterium, *Nitrosomonas europaea*. *Nature Struct. Biol.* **4**, 276–284 (1997).
- Iverson, T. M. et al. Heme packing motifs revealed by the crystal structure of the tetraheme cytochrome c₅₅₄ from *Nitrosomonas europaea*. *Nature Struct. Biol.* **5**, 1005–1012 (1998).
- Otinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1996).
- Collaborative Computational Project No. 4. The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- La Fortelle, E. D., Irwin, J. J. & Bricogne, G. SHARP: A maximum-likelihood heavy-atom parameter refinement and phasing program for the MIR and MAD methods. *Crystalllogr. Comput.* **7**, 1–9 (1997).
- Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F₁-ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
- Brünger, A. T. X-PLOR Version 3.1. *A system for Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
- Engh, R. A. & Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. D* **4**, 392–400 (1991).
- Brünger, A. T. et al. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

Acknowledgements. This work was supported by Volkswagenstiftung (P.K.), Deutsche Forschungsgemeinschaft (P.K.), Fonds der Chemischen Industrie (P.K.), EU-Biotech Project (O.E., A.M., R.H.) and the European Network MASIMO in Human Capital and Mobility (P.K., A.M.). We thank N. Pfennig and H. Beinert who initiated this research, and K. Sulger for assistance in the purification of nitrite reductase.

Correspondence should be addressed to P.K. or O.E. (e-mail: einsle@biochem.mpg.de). The nucleotide sequence of the *S. deleyianum* *nrfa* gene has been deposited with the EMBL Nucleotide Sequence Database (accession no AJ133037). Coordinates of the crystal structure are available from the Protein Data Base (entry code 1QDB).

Proton translocation by cytochrome c oxidase

Michael I. Verkhovsky, Audrius Jasaitis, Marina L. Verkhovskaya, Joel E. Morgan & Mårten Wikström

Helsinki Bioenergetics Group, Department of Medical Chemistry, Institute of Biomedical Sciences and Biocentrum Helsinki, P. O. Box 8, 00014 University of Helsinki, Helsinki, Finland

Cell respiration in mitochondria and some bacteria is catalysed by cytochrome c oxidase, which reduces O₂ to water, coupled with translocation of four protons across the mitochondrial or bacterial membrane^{1–3}. The enzyme's catalytic cycle consists of a reductive phase, in which the oxidized enzyme receives electrons from cytochrome c, and an oxidative phase, in which the reduced enzyme is oxidized by O₂. Previous studies indicated that proton translocation is coupled energetically only to the oxidative phase⁴, but this has been challenged⁵. Here, with the purified enzyme in lipid vesicles, we report time-resolved measurements of membrane potential, which show that half of the electrical charges due to proton-pumping actually cross the membrane during reduction after a preceding oxidative phase. pH measurements confirm that proton translocation also occurs during reduction, but only when immediately preceded by an oxidative phase. We conclude that all the energy for proton translocation is conserved in the enzyme during its oxidation by O₂. One half of it is utilized for proton-pumping during oxidation, but the other half is unlatched for this purpose only during re-reduction of the enzyme.

After the discovery that cytochrome c oxidase works as a proton pump¹, it became important to elucidate the molecular mechanism of proton translocation, and the way it is linked to the reduction of O₂ to water, catalysed by the enzyme. The enzyme's catalytic cycle can be divided into two halves (Fig. 1a). In the oxidative phase, the reduced enzyme (R), which has four electrons in its haem and copper redox centres (Fig. 1b), reacts with O₂ to produce the fully oxidized enzyme (O) and water. This phase involves the intermediate states P and F of the enzyme's haem a₃-Cu_B centre, where the O₂ molecule binds and is reduced to water⁶. In the reductive phase, these redox centres are re-reduced by cytochrome c before another O₂ molecule can bind to initiate the next catalytic cycle.

Equilibrium titrations in mitochondria⁴ indicate that the oxidative phase can be driven backwards from state O to state P by an electrochemical proton gradient, and that all proton translocation by the enzyme is energetically coupled to the P → F and F → O reactions. However, these results also led to the paradigm (see ref. 7) that all proton translocation by the enzyme takes place during the oxidative phase, which is a simplified interpretation of the equilibrium data. These data have been criticized and a model proposed in which one proton is translocated mechanistically and energetically coupled to the reductive phase of the catalytic cycle⁵. However, only direct kinetic determination of proton translocation can definitely establish when it occurs. Here, we report such measurements, which show that all proton translocation does not take place in synchrony with those reaction steps of the oxidative phase that provide the energy for the process. Instead, the protein can conserve part of the

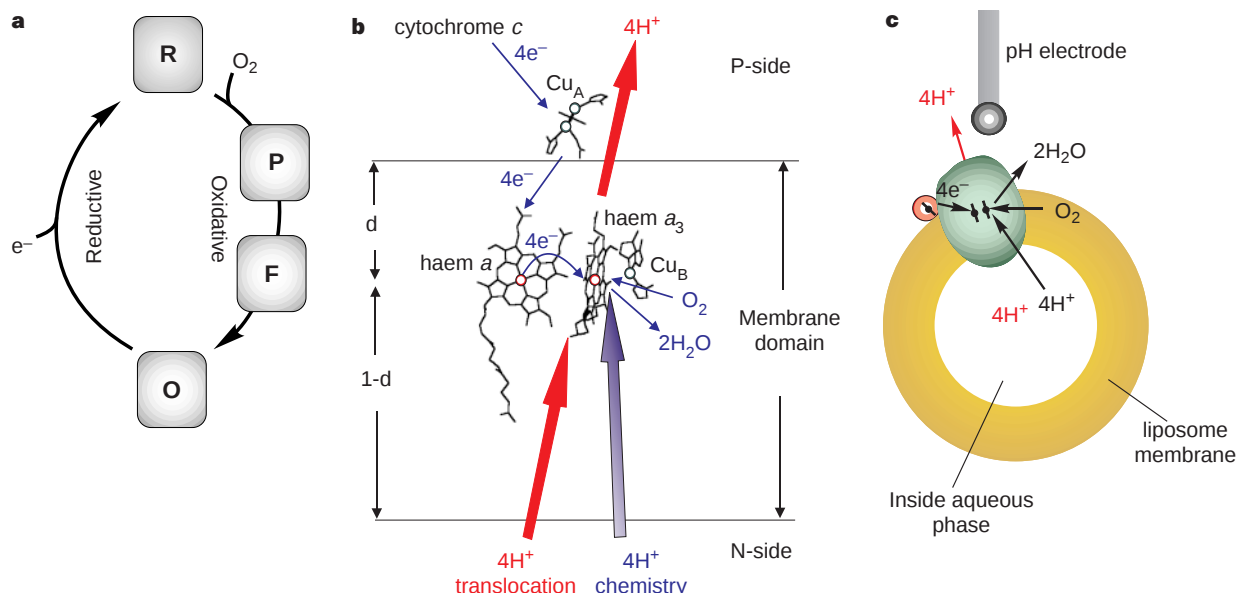


Figure 1 Cytochrome *c* oxidase **a**, The catalytic cycle. In the oxidative phase, the reduced enzyme (R) reacts with O_2 and becomes oxidized to state O through the 'peroxy' (P) and 'ferry' (F) intermediate states of the binuclear haem-copper centre. In the reductive phase, the oxidized state O receives electrons from cytochrome *c* to become re-reduced. **b**, Scheme of cytochrome *c* oxidase in the membrane. Three of the enzyme's redox centres (haem *a*, haem *a*₃ and Cu_B) are embedded in the membrane at a similar depth⁵, which is expressed as the parameter *d*, the fractional dielectric depth from the positively charged P-side⁸. Electrons enter the Cu_A centre from cytochrome *c* on the P-side of the membrane,

and are transferred to haem *a* across **d**, and then to the binuclear haem *a*₃- Cu_B centre of O_2 reduction. Reduction of one O_2 to water by four electrons (blue) is accompanied by uptake of four 'chemical' protons (blue) from the N-side of the membrane across $1 - d$, and by translocation (pumping) of four protons (red) from the N- to the P-side. **c**, Cytochrome *c* oxidase proteoliposome. Cytochrome *c* oxidase (green), inlaid anisotropically in the phospholipid membrane (yellow) of the liposome, receives electrons from cytochrome *c* (red) to catalyse the reduction of O_2 to water, which is coupled to proton translocation (red arrow).

energy so that some proton translocation events are delayed to the reductive phase of the catalytic cycle.

Proton translocation by cytochrome *c* oxidase was determined for the two halves of the catalytic cycle (Fig. 1a), by directly measuring pH changes in a suspension of proteoliposomes (Fig. 1c) inlaid with cytochrome *c* oxidase from bovine heart mitochondria^{8,9}. When the fully reduced enzyme is oxidized by O_2 there is release of only about 1–2 protons (Fig. 2a), in agreement with earlier data¹⁰. The absorbance trace of the lower panel confirms that the entire enzyme population is rapidly oxidized. In contrast, when oxidation is followed by immediate re-reduction, a total of four protons are ejected owing to complete proton-pumping (Fig. 2b). Full re-reduction of the enzyme is not necessary to achieve ejection of four protons, which occurs already after about 40% reduction of the two haem groups, the redox states of which are monitored in the lower trace of Fig. 2b. Preliminary data indicate that input of only one electron after the oxidative phase may be sufficient. On the other hand, reduction of the enzyme in state O, without a preceding oxidative phase, does not lead to significant proton release (data not shown). Thus, proton pumping is incomplete for both the oxidative and reduction phases when assayed alone, and translocation of four protons occurs only if oxidation is immediately followed by re-reduction.

To complement the measurements of proton release, we determined the movement of electrical charge equivalents (*q*) across the membrane during the oxidative and reductive phases. Electrical charge can move across the dielectric by three different means; electron input, proton uptake and events involved in proton translocation (Fig. 1b). Development of membrane potential was measured electrometrically^{9,11}, starting from the reduced CO-bound enzyme in the presence of excess CO and a very low concentration of O_2 . After photolysis of the enzyme-CO complex, only a small fraction of the enzyme (here ~2%) reacts with O_2 and enters the

oxidative phase (Fig. 3a); the majority recombines with CO. The enzyme fraction that is oxidized may be re-reduced immediately at a rate that depends on the concentration of ferrocyanide present as a reductant. After reduction, most of this fraction recombines with CO. Although 2% of it may again react with the O_2 present, this represents only 0.04 per cent of the total enzyme and can be neglected. Therefore, this experimental arrangement allows time-resolved measurement of charge translocation during a single cycle of enzyme oxidation and re-reduction (Fig. 3a). Special care was taken to ensure that full re-reduction of the enzyme occurred at the highest cytochrome *c* concentrations employed.

Figure 3b shows the result of such measurements. Increasing the concentration of reductant (traces 1–4) increases the rate and extent of the slower *reductive* phase to its maximum, where its amplitude accounts for 56.3% (± 3.01 s.d.; 7 independent experiments) of the overall charge translocation. Of the charges, 43.7% cross the membrane during the fast oxidative phase, which is distinguished by the fact that its amplitude is independent of the cytochrome *c* concentration, but varies with the ratio of O_2 to CO. Because full proton-pumping occurs on oxidation with re-reduction (Fig. 2b), whereby a total of eight electrical charges cross the membrane (Fig. 1b, c), it follows that the oxidative and reductive phases contribute to this by 3.5 and 4.5 translocated charge equivalents, respectively.

Electrometric observations⁹ have shown that the oxidative phase is associated with translocation of a total of $10.9d$ charge equivalents across the membrane, where *d* denotes the fractional depth of the haem groups in the membrane dielectric (Fig. 1b). From the finding that the oxidative phase corresponds to translocation of $3.5q$ it therefore follows that *d* equals 0.32 (± 0.02 s.d.), which is in much better agreement with the crystal structure⁵ than the value of ~0.5 originally obtained under equilibrium conditions in mitochondria¹². The larger value for *d* may be the result of

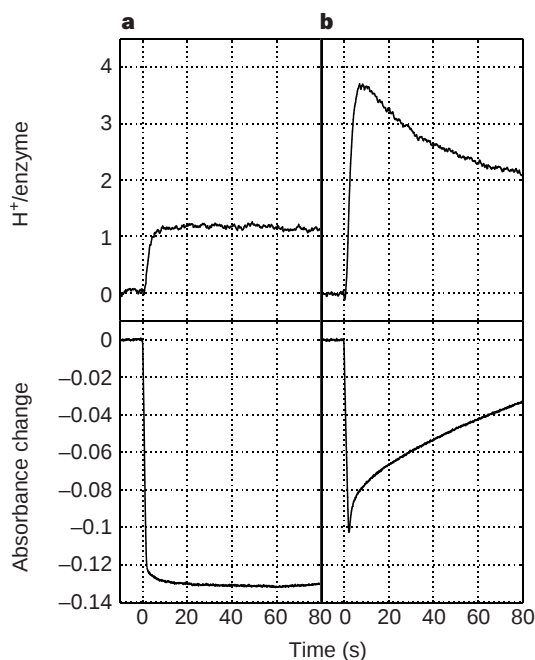


Figure 2 Proton translocation in cytochrome *c* oxidase vesicles. Simultaneous recording of pH (top) and absorbance of haems *a* and *a*₃ at 445–470 nm (bottom). Proteoliposomes at 1.25 μ M enzyme were suspended in 1.2 ml anaerobic 0.1 M KCl medium with 4 μ M valinomycin, pH 7.9, 25 °C. The enzyme was reduced by several 0.1 μ l additions of anaerobic 10 mM hexaammineruthenium(II). **a**, Without ferrocyanochrome *c*; **b**, 1.2 μ M ferrocyanochrome *c* was also present. At zero time 1.25 μ M O₂ was added as a small volume of pure water to start the reaction. A decrease in absorbance at 445–470 nm (downward deflection) indicates oxidation of the haem groups *a* and *a*₃ with no contribution from cytochrome *c*.

secondary polarization effects in the enzyme structure, which do not have time to occur in measurements on the micro- to milli-second time scale.

Half of the eight charges translocated per O₂ reduced are due to the electrical events associated with proton-pumping^{1,6}. The other four charges are due to electron transfer from cytochrome *c* across **d** into the enzyme, and proton uptake from the opposite side of the membrane (across **1** – **d**) to form water at the bimetallic haem-copper site (Fig. 1b, blue arrows). It is possible to estimate the number of charges translocated across the dielectric, relative to **d**, due to these latter events (excluding proton pumping). During oxidation of the fully reduced enzyme, only one electron (from Cu_A) traverses **d** (the other three are already in the haem groups and in Cu_B), and two protons are taken up from the opposite side¹³ (that is, two charges cross **1** – **d**), which yields 2 – **d** charges translocated. During reduction of the oxidized enzyme, three electrons traverse **d** (to reduce the haems and Cu_B), again with uptake of two protons from the opposite side¹⁴, yielding 2 + **d** charge equivalents translocated. From the known value of **d** it then follows that 1.68 q and 2.32 q are translocated because of these ‘chemical’ reactions (Fig. 1b) of the oxidative and reductive phases, respectively. Subtracting this from the total charge translocation in the oxidative (3.5 q) and reductive (4.5 q) phases leads to the conclusion that the proton-pumping events in these phases account for translocation of 1.8 (3.5 – 1.68) and 2.2 (4.5 – 2.33) charges, respectively.

We conclude that when the reduced enzyme is oxidized by O₂, and immediately re-reduced, about two electrical charges are translocated owing to proton pumping events in each phase. The simplified interpretation of the earlier thermodynamic data⁴, that all proton translocation takes place during the oxidative phase, is

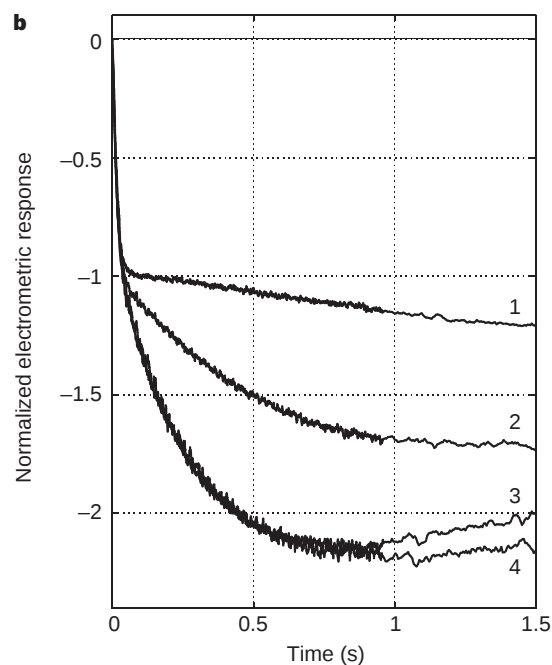
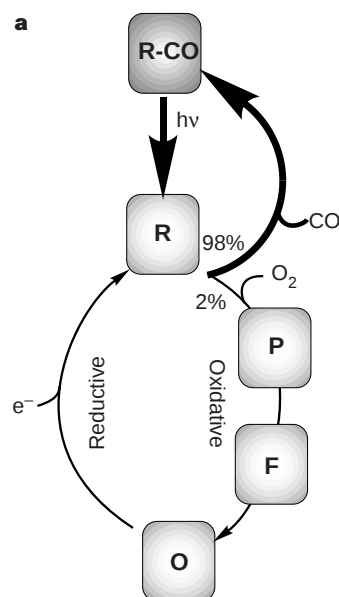


Figure 3 Charge translocation in cytochrome *c* oxidase vesicles. **a**, Scheme of the experiment (see text). **b**, Time-resolved development of membrane potential: 200 mM 1,3-bis[tris(hydroxymethyl)-methylamino]propane buffer with 1 mM CO, 50 mM glucose, 130 μ g ml⁻¹ glucose oxidase, 50 μ g ml⁻¹ catalase, pH 7, 25 °C, and cytochrome *c* oxidase proteoliposomes attached to the measuring membrane⁹. [O₂] was estimated to be 10 nM. Reduced cytochrome *c* was present at 20 (trace 1), 70 (trace 2) and 170 (trace 3) μ M. In trace 4, 170 μ M ferrocyanochrome *c* was supplemented with 10 μ M N,N,N',N'-tetramethyl-*p*-phenylenediamine (TMPD) to ensure full reduction of cytochrome *c* and the oxidase. All curves are normalized to the amplitude and rate of the fast phase. Each curve is the average of 16 flashes.

thus erroneous. However, our data also show that proton translocation is not coupled energetically to the reductive phase, as proposed⁵, because in that case it should occur during enzyme reduction without prior oxidation. This is supported by the fact that reduction of the enzyme's redox centres (with $E_{m,7}$ values of 0.2 – 0.4 V) by cytochrome *c* ($E_{m,7} \approx 0.25$ V) would hardly yield sufficient energy for proton translocation, in addition to the energy required for electron input and proton uptake ('chemistry', Fig. 1b).

As proton pumping during reduction occurs only when immediately preceded by an oxidative phase, all the energy for proton translocation may indeed be conserved during the oxidative phase⁴. Half of this energy is converted into electrogenic proton movement across the dielectric during oxidation, but the other half must be conserved in the enzyme species O (Fig. 1a), which is the product of the oxidative phase. This energy is unlatched for proton translocation only if the enzyme receives electrons immediately; otherwise it is lost as heat.

A primary result from our work is the idea that energy from the redox processes in the oxidative phase of the catalytic cycle can be conserved in a metastable state of the enzyme. The bimetallic haem-copper centre is in a ferric/cupric form^{4,6} in state O. Our results thus indicate that there may be two forms of this state, O and O~, of which O~ is the immediate product of the oxidative phase with energy conserved for proton translocation. The time constant for dissipation of O~ to O is yet to be accurately determined, but it is less than 20 s under the conditions reported here, and might therefore be related to the previously described relaxation process of the binuclear haem a_3 /Cu_B site¹⁵.

It may be of historical interest to note that the proposed state O~ is reminiscent of the concept of energized protons in membranes advocated by Williams¹⁶, and analogous to high-energy forms of cytochromes postulated in Slater's¹⁷ chemical hypothesis of oxidative phosphorylation. Proton translocation by cytochrome *c* oxidase must, in part, be considered to be indirectly coupled to the O₂ reduction chemistry. The protein structure is used as a temporary store of energy. Like a spring, it is stretched by the redox reactions of the oxidative phase. When triggered by electron input, the spring relaxes and drives proton translocation. □

Received 8 April; accepted 17 May 1999.

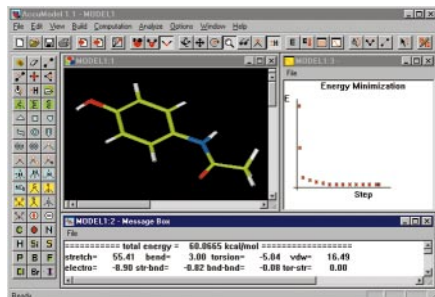
1. Wikström, M. Proton pump coupled to cytochrome *c* oxidase in mitochondria. *Nature* **266**, 271–273 (1977).
2. Solioz, M., Carafoli, E. & Ludwig, B. The cytochrome *c* oxidase of *Paracoccus denitrificans* pumps protons in a reconstituted system. *J. Biol. Chem.* **257**, 1579–1582 (1982).
3. Puustinen, A., Finel, M., Virkki, M. & Wikström, M. Cytochrome *o*(bo) is a proton pump in *Paracoccus denitrificans* and *Escherichia coli*. *FEBS Lett.* **249**, 163–169 (1989).
4. Wikström, M. Identification of the electron transfers in cytochrome oxidase that are coupled to proton-pumping. *Nature* **338**, 776–778 (1989).
5. Michel, H. The mechanism of proton pumping by cytochrome *c* oxidase. *Proc. Natl Acad. Sci. USA* **95**, 12819–12824 (1998).
6. Babcock, G. T. & Wikström, M. Oxygen activation and the conservation of energy in cell respiration. *Nature* **356**, 301–309 (1992).
7. Gennis, R. B. How does cytochrome oxidase pump protons? *Proc. Natl Acad. Sci. USA* **95**, 12747–12749 (1998).
8. Verkhovsky, M. I., Belevich, N., Morgan, J. E. & Wikström, M. Proton linkage of cytochrome *a* oxidoreduction and proton translocation by the haem-copper oxidases. *Biochim. Biophys. Acta* (in the press).
9. Jasaitis, A., Verkhovsky, M. I., Morgan, J. E., Verkhovskaya, M. L. & Wikström, M. Assignment and charge translocation stoichiometries of the electrogenic phases in the reaction of cytochrome *c* with dioxygen. *Biochemistry* **38**, 2697–2706 (1999).
10. Oliveberg, M., Hallén, S. & Nilsson, T. Uptake and release of protons during the reaction between cytochrome *c* oxidase and molecular oxygen: A flow-flash investigation. *Biochemistry* **30**, 436–440 (1991).
11. Drachev, L. A. *et al.* Direct measurement of electric current generation by cytochrome oxidase, H⁺-ATPase and bacteriorhodopsin. *Nature* **249**, 321–324 (1974).
12. Hinkle, P. & Mitchell, P. Effect of membrane potential on the redox poise between cytochrome *a* and cytochrome *c* in rat liver mitochondria. *Bioenergetics* **1**, 45–60 (1970).
13. Mitchell, R., Mitchell, P. & Rich, P. R. Protonation states of the catalytic intermediates of cytochrome *c* oxidase. *Biochim. Biophys. Acta* **101**, 188–191 (1992).
14. Mitchell, R. & Rich, P. R. Proton uptake by cytochrome *c* oxidase on reduction and on ligand binding. *Biochim. Biophys. Acta* **1186**, 19–26 (1994).
15. Armstrong, F., Shaw, R. W. & Beinert, H. Cytochrome *c* oxidase. Time dependence of optical and EPR spectral changes related to the 'oxygen-pulsed' form. *Biochim. Biophys. Acta* **722**, 61–71 (1983).
16. Williams, R. J. P. Possible functions of chains of catalysts. *J. Theoret. Biol.* **1**, 1–17 (1961).
17. Slater, E. C. Mechanism of phosphorylation in the respiratory chain. *Nature* **172**, 975–978 (1953).

Acknowledgements. We thank the Academy of Finland, the University of Helsinki, and the Sigrid Jusélius Foundation for financial support, E. Haasanen for technical assistance and T. Aalto for glassware design.

Correspondence and requests for materials should be addressed to M.W. (e-mail: Marten.Wikstrom@Helsinki.Fi).

Super models in the laboratory

A culling of molecular modelling and other chemistry programs from commercial and .edu sources. Packages vary from the educational to the 'professional', with some with a foot in both camps.



AccuModel: calculations without the expense.

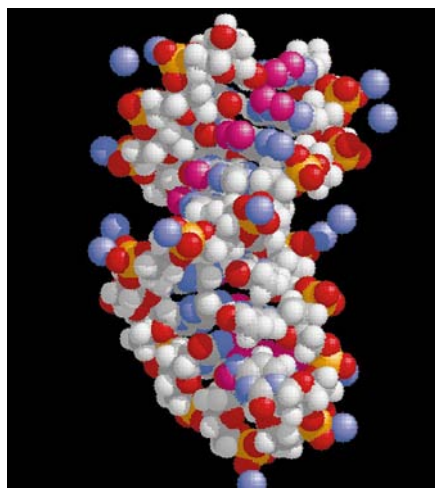
AccuModel

From Microsimulations
www.microsimulations.com

A new 32-bit release

AccuModel gives users several ways to create models. Using a 2D drawing program, you can import structures in several file formats. ISIS/Draw and ISIS/Base users can automatically convert 2D structures into 3D models by clicking the AccuModel add-in button on the ISIS menu bar. The AccuModel tool palette can be used to assemble 3D models just as you would with plastic models. AccuModel also reads PDB and other 3D structure files. The manufacturers say that the combination of the MM3 and UFF force fields allows AccuModel to perform calculations previously only available on expensive workstations. Models can be displayed in space-filling, ball-and-stick or wireframe formats, displaying four different views at one time. The models can be transferred to word processors or graphics programs using copy-and-paste, or exported in several different file formats for use in other chemistry software.

Reader Service No. 100



DNA, by Thomas Jefferson University's AMMP.

AMMP

From Thomas Jefferson University
asterix.jcl.tju.edu

It stands for 'Another Molecular Mechanics Program', really

AMMP is a fully-featured molecular mechanics and dynamics program. It can run fast long-range electrostatics and non-bonded terms. AMMP runs without cutoffs in times comparable to or better than an 8-Å cutoff in other programs. The program is claimed to be stable with difficult problems such as parallel-stranded DNA. In addition to standard formulations, AMMP has non-point charges, explicit Debye screening, accurate polarization models, non-harmonic bond and angle formulations, and coupled bond-angle terms available. AMMP is extensively vectorized and runs under Unix, Windows 3.1 and Windows 95. As an example of an earlier release of AMMP at work, see R. W. Harrison and I. V. Kurinov, 'Prediction of novel serine protease inhibitors' in *Nature Structural Biology* 1, 735-743; 1994).

Reader Service No. 101

ChemSite

From Pyramid Learning www.chemsite1.com

A chemical modelling and visualization program

ChemSite features sophisticated 3-D graphics and molecular dynamics animation. This program is recommended as a teaching tool, showing students how molecules may be easily constructed and viewed in the 3D virtual-reality world of the PC. Real-time rotation with the mouse in full 3D perspective conveys molecular structure effectively, and ray-traced space-filling CPK models give a sense of the size and shape of molecules. Complicated polymers such as proteins and DNA/RNA molecules can readily be shown as 3D ribbons and cylinders for visualization of secondary and tertiary structure. For lectures and publications, ChemSite can export images to any graphics-capable word processor or publishing software.

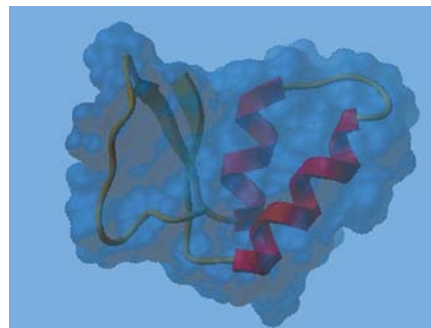
Reader Service No. 102

ICM and ICMlite

From Molsoft <http://molsoft.com>

A modular package for molecular modelling

The ICM package consists of the main program and may include any single or a combination of three optional modules: ICM-



The crambin molecule, combined view by ICM.

REBEL, ICM-bioinformatics and ICM-dock. Key features of the main program include a shell for molecules, numbers, strings, vectors, matrices, tables and sequences; graphics and stereo; imaging and vectorized postscript; advanced animation and movies; mathematics, statistics and plotting; and tools for CGI development. Superpositions, structural alignment and Ramachandran plots are amongst the functions. An offer running now, on <http://216.103.7.98/download.htm>, provides ICMlite free to academics in versions for SGI, DEC, SunOS (Solaris), Windows 95/NT and Linux (i386 and PowerPC).

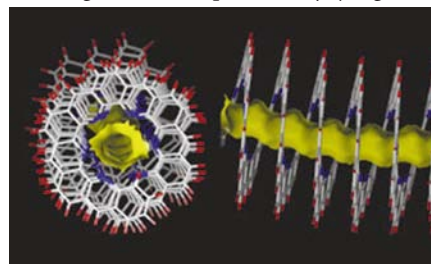
Reader Service No. 103

SYBYL 6.5.2

From Tripos www.tripos.com

The latest release of a tool optimized for drug discovery

SYBYL is a comprehensive computational tool kit for molecular design and analysis, with a special focus on the creation of new chemical entities. SYBYL provides the construction and analysis tools for both organic and inorganic molecular structures. SYBYL 6.5 introduced the ability to generate MOLCAD surface displays of the regions where molecular surfaces are in proximity, colour-coded to show the distances between the surfaces. A ProTable enhancement shows colour-coded Ramachandran maps, allowing users to qualitatively judge the



SYBYL: looks at the backbone.

quality of the protein backbone conformation. Latest version is now SYBYL 6.5.2, which takes account of the fact that the Protein Data Bank adopted Gnu compression (gzip, .gz suffix) in October last year.
Reader Service No. 104

TINKER 3.7

From Washington University

<http://dasher.wustl.edu/tinker>

A set of small programs adding up to a molecular modelling package

TINKER molecular modelling software is a complete and general package for molecular mechanics and dynamics, with some special features for polypeptides. TINKER can use any of several common parameter sets, such as AMBER-95, AMBER/OPLS, CHARMM-22, MM2(91), MM3(96), OPLS All-Atom, ENCAD, MMFF-94 and the TINKER polarizable multipole parameter set. Version 3.7 was released on 16 June 1999. TINKER, with full source code, is available free of charge.

Reader Service No. 105

Binding Site Analysis

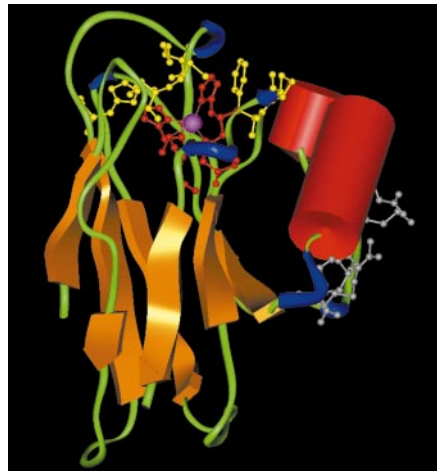
From Molecular Simulations, Inc.

www.msi.com

Tools for structure-based drug discovery

Insight II is a 3D graphical environment for molecular modelling. A recent addition to its armoury is the Binding Site Analysis program. This combines several tools for the identification and characterization of a protein's binding site, then uses those characteristics to look for similar features in other proteins of known structure. The Active Site Finding tool automatically locates binding sites on the surface, the Evolutionary Trace Tool analyses a multiple sequence alignment to identify functionally conserved residues, and the Template Based Searching function is used to develop a structural motif based on the characterization of the binding site. Insight II runs on Silicon Graphics and IBM UNIX workstations and servers.

Reader Service No. 106



Insight II aims at high-end users.



ChemDraw: now at 5.0 or higher.

ChemDraw Ultra 5.0

From Camsoft

www.chemdraw.com

Evolution of a popular package

Latest version of what the makers call the "world's most popular chemical drawing package", is ChemDraw Ultra 5.0. New features include a name-to-structure generator (Name=Struct), built-in structure naming via Beilstein's AutoNom Version 2.1 program, estimated NMR line spectra display, spectral file viewing, a new version of the ChemDraw Plugin, and improved programming interfaces.

Reader Service No. 107

AutoNom

From Beilstein

www.beilstein.com

For AutoNom read AUTOMATIC NOMenclature

Version 4.0 of the AutoNom program from Beilstein Informationssysteme GmbH introduces chiral interpretation (*R*, *S* configurations) and double-bond topology (*E*, *Z* descriptors) into a program that makes use of IUPAC nomenclature rules to generate IUPAC names automatically from chemical structures. Also new in this release is support for CAS ring-system naming conventions. AutoNom 4.0 runs in either Windows or Macintosh operating systems.

Reader Service No. 108

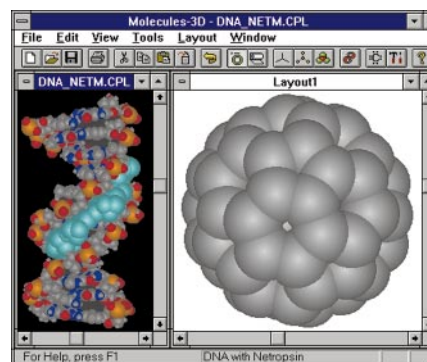
Molecules 3D and 3DPro

From Molecular Arts Corp

www.molecules.com

Molecular modelling for teaching and research use

The Molecules-3D package enables users to build, rotate and examine 3D molecular models. With Molecules-3D, you can transfer 3D models and sophisticated 3D graphic images into other applications. For professionals and advanced students, Molecules-3D Pro (now at version 2.6) includes all the features of Molecules-3D. In addition, Molecules-3D Pro includes a larger struc-



Mol3D, a dual-purpose approach.

ture library (over 2,000 models), the ability to work with larger models and display multiple structures, and greater control over atom parameters and energy minimization.

Reader Service No. 109

HyperChem5

From Hypercube, Inc.

www.hyper.com

PC-based molecular modelling, with a choice of functionality

HyperChem 5.1 is available in three forms, depending on the functionality required. At entry-level, Hyper-Chem Standard lacks the ability to interface to outside programs such as ChemPlus, Q-Chem, or the Chemist's Developer Kit (CDK). With HyperChem Professional CDK is included, a set of tools that enables chemists to quickly construct an interface between HyperChem and other chemical software programs. It also offers customization and a full scripting language for advanced users. HyperChem Suite includes HyperChem Professional plus a new release of ChemPlus Release 5.1.

Reader Service No. 110

These notes are compiled in the Nature office from information provided by the manufacturers or distributors. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENT

Silver Stain Kit II for Electrophoresis

**Rapid, sensitive detection
of proteins & nucleic acids**

Wako

Wako BioProducts
Wako Chemicals USA, Inc.
1600 Bellwood Road
Richmond, VA 23237
Telephone: (804) 271-7677
Facsimile: (804) 271-7791
(800) 992-9256
bioproducts@wakousa.com

READER ENQUIRY NO. 41